

Query Match	31.58;	Score 576.4;	DB 6;	Length 1914;
Best Local Similarity	99.8%;	Pred. No. 3.4e-83;		
Matches 577; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

125	TTCAACATCATGTGAATATAGTGGCCCAATGAGAACAAAGCTTGCAATGTTCCAAAG	184
1337	TTCAACATCATGTGAATATAGTGGCCCAATGAGAACAAAGCTTGCAATGTTCCAAAG	1399
185	AAGGACGATACACTTTGTAGGATATGTGCTGAAGTTCTTCGGAAACTTGCAGCT	244
1397	AAGGACGATACACTTTGTAGGATATGTGCTGAAGTTCTTCGGAAACTTGCAGCT	1455
245	TTGGATTAGAAAGACTCTGTGGCTCTCACTTGGAAATTCACAAAGTGGTTGCACCTC	304
1457	TTGGATTAGAAAGACTCTGTGGCTCTCACTTGGAAATTCACAAAGTGGTTGCACCTC	1511
305	CTCATTAAGAACATGTTTGCACCTGTGCGAAGGCTTTCGTATGCAATACATCTGCTTC	364
1517	CTCATTAAGAACATGTTTGCACCTGTGCGAAGGCTTTCGTATGCAATACATCTGCTTC	1577
365	TTGGTTTATCAGATTATTTCTTTCTTAAAGCTCCCTGAAGGGCAATATCATTTGGCTTG	424
1577	TTGGTTTATCAGATTATTTCTTTCTTAAAGCTCCCTGAAGGGCAATATCATTTGGCTTG	1638
425	GGGTATATAGTATTACTTATTTGATCTTGCAGGCAATATCAAATTAATCTCCACATT	484
1637	GGGTATATAGTATTACTTATTTGATCTTGCAGGCAATATCAAATTAATCTCCACATT	1698
485	TCCAGTGAACAGATGTACATTAAGATTAAGATGACGTGGCTATTTGGTTGAAGGGATTA	544
1697	TCCAGTGAACAGATGTACATTAAGATTAAGATGACGTGGCTATTTGGTTGAAGGGATTA	1758
545	CAGAGCCCAATTAAGATTTAAATATATTCATTAATATTTTATTTGGAAGGTGGCTGG	604
1757	CAGAGCCCAATTAAGATTTAAATATATTCATTAATATTTTATTTGGAAGGTGGCTGG	1818
605	AGAGAGCTGAGGATTTCCAGACCTTGTAAAGTTTATTTGGGAGAACATAGGCCAAT	664
1817	AGAGAGCTGAGGATTTCCAGACCTTGTAAAGTTTATTTGGGAGAACATAGGCCAAT	1878
665	AATCATGACCTCTTCCAGGCAATTTTATAGACAGATGTC 702	
1877	AATCATGACCTCTTCCAGGCAATTTTATAGACAGATGTC 1914	

SULT 6				
CUS				
FINITION	AX302888	381 bp	DNA	linear
CESSION	Sequence 208 from Patent WO01/9286,			
RSION	AX302888			
YWORDS	AX302888.1	GI:17383386		
URCE				
ORGANISM	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
ERENCE	1			
AUTHORS	Jiang, Y., Dillon, D.C., Mitcham, J.L., Xu, J., Hatlocker, S.L. and			
	Hepler, W.T.			
ITILE	Compositions, and methods for the therapy and diagnosis of breast			
	Cancer			
JOURNAL	Patent: WO 01/9286-A 208 25-OCT-2001;			
	CORIXA CORPORATION (US)			
ATURES	Location/Qualifiers			
	1. 381			
source	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			

BASE COUNT	106 a	68 c	51 g	142 t	14 others
ORIGIN					

Query Match	19.7%	Score 360.6;	DB 6;	Length 381;
Best Local Similarity	95.3%	Pred. No. 2.2e-48;		
Matches 363; Conservative	0;	Mismatches 18;	Indels 0;	Gaps 0;

QY	949	AGAGGAACATTTTCCATGCGCTGAGTTTTCCTATCCACCCCTAACCTAACTAATTA	1008
Db	1	AGAGGAATATTTTCCATGCGCTGATATCTTTCCTATCCACCCCTAACCTAATTA	60
QY	1009	CTTAGTCGTCTTGTATTAAGCAAGATATTACCTTTAACTTGCGCTCTTACTCTTGGCCTT	1066
Db	61	CTTAGTCGTCTTGTATTAAGCAAGATATTACCTTTAACTTGCGCTCTTACTCTTGGCCTT	120
QY	1069	TAGCTACTAATAAAGTTTGATATAGGCAATTATTAATATTTGAGTCATTCATGGTAT	1128
Db	121	TAGCTAATCTAATAAAGTTTGATATAGGCAATTATTAATATTTGAGTCATTCATGGTAT	180
QY	1129	CTCTCATGTTGATGATATTTTCAACATAGATCATATGTTTTCCTCAGAGTCC	1188
Db	181	CTCTCATGTTGATGATATTTTCAACATAGATCATATGTTTTCCTCAGAGTCC	240
QY	1189	ATTAAATCATTTATTTCCCTTACTTCTCACCTCGTGTGAACAATTTGAACAAGTGATTT	1248
Db	241	ATTAAATCATTTATTTCCCTTACTTCTCACCTCGTGTGAACAATTTGAACAAGTGATTT	300
QY	1249	GGGAACCCATTTTGGAAAAACAGATTCATAGTCATGAAAAATGGAACCTTCATATTTCTG	1308
Db	301	GGGAACCCATTTTGGAAAAACAGATTCATAGTCATGAAAAATGGAACCTTCATATTTCTG	360
QY	1309	TTTTTGAAGAATGTGGCCAT	1329
Db	361	TTTTTGAAGAATGTGGCCAT	381

RESULT 7					
D64141	D64141	3090 bp	mRNA	linear	ROD 13-FEB-1999
LOCUS					
DEFINITION	Mus musculus mRNA for protein-tyrosine-phosphatase,				complete cds.
ACCESSION	D64141				
VERSION	D64141.1 GI:2665457				
KEYWORDS	protein-tyrosine-phosphatase.				
SOURCE	Mus musculus testis cDNA to mRNA.				
ORGANISM	Mus musculus				

REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE
1 (sites) Ohsugi, M., Kuramochi, S., Matsuda, S. and Yamamoto, T. Molecular cloning and characterization of a novel cytoplasmic protein-tyrosine phosphatase that is specifically expressed in spermatocytes J. Biol. Chem. 272 (52), 33092-33099 (1997)	2 (bases 1 to 3050) Ohsugi, M. Direct Submission Submitted (15-SEP-1995) Miho Ohsugi, The University of Tokyo, Department of Oncology, 4-6-1 Shirokanebashi, Minato-ku, Tokyo 108, Japan (E-mail: mohsugi@elms.u-tokyo.ac.jp, Tel: 03-5449-5305, Fax: 03-5449-5413)

```

FEATURES
Source
Location/Qualifiers
1..3090
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="testis"
218..1498
/EC_number="3.1.3.48"
/codon_start=1
/product="protein-tyrosine-phosphatase"
/protein_id="BA23761.1"
/db_xref="GI:2655458"
/translation="MSPKRVGRKGTGHNDNEEGNSGNLNLRLNSLPSSSKMPTPKPI

```



RESULT 2  
ID 055082 PRELIMINARY; PRT; 426 AA.  
AC 055082;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DE Protein-tyrosine-phosphatase, 20, last annotation update)  
GN PTPN20.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCB1\_TaxID-10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-TESTIS;  
RX MEDLINE-98070510; PubMed-9407093;  
RA Ohnogi M., Kuramochi S., Matsuda S., Yamamoto T.;  
RT "Molecular cloning and characterization of a novel cytoplasmic  
protein-tyrosine phosphatase that is specifically expressed in  
spermatocytes."  
RL J. Biol. Chem. 272:33092-33099 (1997).  
DR EMBL; D64141; BAA23761.1; -  
DR HSP; 006124; 2SHP.  
DR MGD; MGI:1196295; Ptpn20.  
DR InterPro: IPR000387; TYR\_phosphatase.  
DR InterPro: IPR000242; TYR\_PP.  
DR Pfam: PF00102; Y\_phosphatase; 1.  
DR PRINTS; PR00700; PTPPHPTASE.  
DR SMART; SM00194; PTPC; 1.  
DR PROSITE; PS00363; TYR\_PHOSPHATASE\_1; 1.  
DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
KW Hydrolase.  
SO SEQUENCE 426 AA; 4918 MW; 2B35FB13379502F4 CRC64;  
Query Match 32.8%; Score 121; DB 11; Length 426;  
Best Local Similarity 61.1%; Pred. No. 5.6e-07;  
Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
QY 1 MOEFALRLKLNLPGEFNSGNPSNREKRYRDIILPF 36  
DB 166 IREFLELEQMTLPDDFNSGNTLQNRDKNRIDILPY 201  
ID 063476 PRELIMINARY; PRT; 297 AA.  
AC 063476;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
DE Protein tyrosine phosphatase epsilon M precursor (Fragment).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCB1\_TaxID-10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SPLEEN;  
RX MEDLINE-96158952; PubMed-8579581;  
RA Nakamura K., Mizuno Y., Kikuchi K.;  
RT "Molecular cloning of a novel cytoplasmic protein tyrosine phosphatase  
pPpase10n."  
RL Biochem. Biophys. Res. Commun. 218:726-732 (1996).  
DR EMBL; D78613; BAA11433.1; -  
DR HSP; P18052; LYFO.  
DR InterPro: IPR000242; TYR\_PP.  
DR Pfam: PF00102; Y\_phosphatase; 1.  
DR PRINTS; PR00700; PTPPHPTASE.  
DR SMART; SM00194; PTPC; 1.

DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
KW Signal.  
FT SIGNAL. 1 18 POTENTIAL.  
FT NON\_TER 297  
SQ SEQUENCE 297 AA; 34221 MW; BCFMA633BF526DF CRC64;  
Query Match 24.0%; Score 88.5; DB 11; Length 297;  
Best Local Similarity 28.9%; Pred. No. 0.0049;  
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
QY 2 QEFVALRLKLNLPGEFNSGNPSNREKRYRDIILPF 39  
DB 136 EEFNSLPSGHIQGFELANKEENREKRYRDIILPF 139  
ID 08TE48 PRELIMINARY; PRT; 536 AA.  
AC 08TE48;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DE Tyrosine phosphatase epsilon PDI.  
GN PTPRE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCB1\_TaxID-9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC "Characterisation, expression and functional aspects of a novel  
protein tyrosine phosphatase epsilon isoform."  
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ430580; CAD23182.1; -  
SQ SEQUENCE 536 AA; 60799 MW; 06C29CBA8CCDE6A CRC64;  
Query Match 24.0%; Score 88.5; DB 4; Length 536;  
Best Local Similarity 28.9%; Pred. No. 0.0098;  
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
QY 2 QEFVALRLKLNLPGEFNSGNPSNREKRYRDIILPF 39  
DB 79 EEFNSLPSGHIQGFELANKEENREKRYRDIILPF 139  
ID 40 GY-----SGPNERTT--FW 51  
DB 139 GYKKKKFTAAQGPKEVYDFW 161  
RESULT 5  
ID 096K06 PRELIMINARY; PRT; 642 AA.  
AC 096K06;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
DE Tyrosine phosphatase epsilon.  
GN PTPRE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCB1\_TaxID-9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Wabakken T.K., Hauge H., Flinne E.F., Wiedlocha A., Aasheim H.C.;  
RT "Expression of human protein tyrosine phosphatase epsilon in  
leukocytes, a potential ERK-pathway regulating phosphatase."  
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 11, 2003, 11:44:37 ; Search time 211 Seconds

(without alignments)  
715.090 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369  
Sequence: 1 MQEFMALEKMLPGFENSGN.....TTFMHSNGSNVSLLRCA 67

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ .p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09935703/runat\_06062003\_095210\_6144/app\_query.fasta.1.263  
-DB=N-Geneseq.101002 -QFMT=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPT=0  
-LISTEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human4.0.cdi  
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09935703.GCGN.1.1.396 -tunat\_06062003\_095210\_6144 -NCP=6 -ICPU=3  
-NO\_MMAP -IARG=QDERT -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N-Geneseq.101002.\*  
1: /SID52/gcgdata/geneseg/genesegn-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseg/genesegn-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseg/genesegn-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseg/genesegn-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseg/genesegn-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseg/genesegn-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseg/genesegn-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseg/genesegn-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseg/genesegn-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseg/genesegn-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseg/genesegn-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseg/genesegn-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseg/genesegn-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseg/genesegn-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseg/genesegn-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseg/genesegn-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseg/genesegn-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseg/genesegn-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseg/genesegn-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseg/genesegn-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseg/genesegn-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseg/genesegn-emb1/NA2001.DAT.\*  
23: /SID52/gcgdata/geneseg/genesegn-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseg/genesegn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	369	100.0	1832	22	AAH43389	PTPase cDNA clone
2	239.5	64.9	1914	24	AAH24022	Human protein phosphatase
3	231.5	62.7	3264	23	AAH70229	DNA encoding novel
4	123	33.3	1692	20	AAH81746	Mouse PTP05 isofor
5	123	33.3	1785	20	AAH81744	Mouse PTP05 isofor
6	123	33.3	1896	20	AAH81745	Mouse PTP05 isofor
7	88.5	24.0	2160	24	ABK84570	Human cDNA differe
8	88.5	24.0	2389	22	AAH5987	Human poly nucleoti
9	88.5	24.0	2393	22	AAH59773	Human poly nucleoti
10	88.5	24.0	2689	22	AAH98486	Human Est-derived
11	88.5	24.0	5425	24	ABH99311	Mouse Ischaemic co
12	88	23.8	1747	17	AAH35308	Rat M1P7 gene cDN
13	86	23.3	1624	17	AAH35310	Human SH-PTP1 cDN
14	86	23.3	1788	22	AAH26454	SH-1 DNA. Unden
15	86	23.3	1788	22	AAH26466	SH-1 activated do
16	86	23.3	1788	22	AAH26467	SH-1 activated do
17	86	23.3	1788	22	AAH26468	SH-1 activated do
18	86	23.3	1788	22	AAH26469	SH-1 activated do
19	86	23.3	1788	22	AAH26470	SH-1 activated tr
20	86	23.3	1788	22	AAH26471	SH-1 activated tr
21	86	23.3	1788	22	AAH26472	SH-1 activated qu
22	86	23.3	1788	22	AAH26473	SH-1 activated qu
23	86	23.3	1788	22	AAH26474	SH-1 activated qu
24	86	23.3	1788	22	AAH26475	SH-1 activated qu
25	86	23.3	1788	22	AAH26484	SH-1 D59A mutant
26	86	23.3	1788	22	AAH26485	SH-1 E74A mutant
27	86	23.3	1866	21	AAH32917	Human HCP phosphid
28	86	23.3	2076	24	ABK34867	Human cDNA encodin
29	86	23.3	2145	17	AAH35306	Human SH-PTP1 gene
30	86	23.3	2175	24	ABK63571	Rat sequence differe
31	86	23.3	2277	21	AAH09637	Human Src homology
32	86	23.3	2277	24	ABK8611	Human cDNA differe
33	86	23.3	2315	24	ABO5031	Human ovarian anti
34	86	23.3	2472	21	AAH32916	Human HCP/GST fusi
35	86	23.3	2532	19	AAH71075	Green fluorescent
36	86	23.3	2562	19	AAH71076	Protein tyrosine p
37	81.5	22.1	14738	22	AAH02949	Human reproductive
38	81.5	22.1	14738	22	AAH02951	Human reproductive
39	81.5	22.1	222930	24	ABK84349	Human cDNA differe
40	79.5	21.5	2409	17	AAH36882	Human receptor typ
41	79.5	21.5	2409	20	AAH36882	Human receptor typ
42	79.5	21.5	2826	24	ABO5083	Human ovarian anti
43	79.5	21.5	2872	13	AAH021001	Marine receptor ty
44	79.5	21.5	2872	17	AAH36881	Marine receptor ty
45	79.5	21.5	2872	20	AAH29853	Mouse receptor typ

## ALIGNMENTS

RESULT 1  
ID AAH43389 standard; cDNA: 1832 BP.  
XX  
AC AAH43389;  
XX  
DT 15-NOV-2001 (first entry)  
XX  
DE PTPase cDNA clone HATBM23.  
XX  
KW Protein tyrosine phosphatase; PTPase; neural disorder; preservative;  
KW Alzheimer's disease; Parkinson's disease; immune system disorder;  
KW rheumatoid arthritis; Grave's disease; muscular disorder; wound healing;  
KW reproductive disorder; pulmonary disorder; cardiovascular disorder;  
KW arylhydrolase; infectious disease; viral hepatitis; multiple sclerosis;  
KW hyperproliferative disorder; neoplasm; epithelial cell proliferation;  
KW endocrine disorder; diabetes mellitus; transplantation; weight disorder;  
KW hair loss; skin aging; sunburn; food additive; ss.



XX OS Homo sapiens.  
 XX XX  
 XX Key Location/Qualifiers  
 XX FT 19..222  
 XX FT CDS /tag- a  
 XX FT /product- "Prtase"  
 XX  
 XX MO20016706-A1.  
 XX  
 XX PD 13-SEP-2001.  
 XX  
 XX PF 22-FEB-2001; 2001WO-US05496.  
 XX  
 XX PR 03-MAR-2000; 2000US-0186658.  
 XX PR 16-MAR-2000; 2000US-0189881.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX PI Shl Y, Ruben SM;  
 XX  
 XX DR MPI: 2001-565582/63.  
 XX DR P-PSDB; AAB47496.  
 XX  
 XX PT Novel isolated protein tyrosine phosphatase polypeptide useful for  
 XX PT treating and preventing Alzheimer's disease, rheumatoid arthritis,  
 XX PT Grave's disease, arrhythmias, neoplasms, multiple sclerosis and  
 XX PT diabetes mellitus  
 XX  
 XX PS Claim 1; Page 276-77; 286pp; English.  
 XX  
 XX CC The sequences given in AAH43389-93 encode protein tyrosine phosphatase  
 XX CC (Prtase) polypeptides. The Prtase polypeptides, or the cDNA encoding  
 XX CC them, are useful for preventing, treating or ameliorating a medical  
 XX CC condition in a mammalian subject. They are useful for diagnosing,  
 XX CC preventing or treating neural disorders (e.g. Alzheimer's disease,  
 XX CC Parkinson's disease), immune system disorders (e.g. rheumatoid  
 XX CC arthritis, Grave's disease), muscular disorder, reproductive  
 XX CC disorders, pulmonary disorders, cardiovascular disorders  
 XX CC (e.g. arrhythmias), infectious diseases (e.g. viral hepatitis),  
 XX CC hyperproliferative disorders (e.g. neoplasms), diseases of the cellular  
 XX CC level (e.g. multiple sclerosis), endocrine disorders (e.g. diabetes  
 XX CC mellitus), and wound healing and epithelial cell proliferation.  
 XX CC They are also useful for preventing hair loss, to prevent skin aging  
 XX CC due to sunburn, to maintain organs before transplantation, to modulate  
 XX CC mammalian characteristics, to treat weight disorders, to modulate  
 XX CC mammal's mental or physical state, or as a food additive or  
 XX CC preservative.  
 XX  
 XX SO Sequence 1832 BP; 560 A; 299 C; 321 G; 652 T; 0 other;  
 XX  
 XX Alignment Scores:  
 XX Pred. No.: 1,94e-41 Length: 1832  
 XX Score: 369.00 Matches: 67  
 XX Percent Similarity: 100.008 Conservative: 0  
 XX Best Local Similarity: 100.008 Mismatch: 0  
 XX Query Match: 100.008 Indels: 0  
 XX DB: 22 Gaps: 0  
 XX  
 XX US-09-935-703-7 (1-67) x AAH43389 (1-1832)  
 XX  
 XX OY 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20  
 XX DB 19 ATGCGAGAAATTATGCTTTAGAACTTAAAGATCTGCTGTGATCAACTCTGGGAAAT 78  
 XX OY 21 GlnPrtSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHisGly 40  
 XX DB 79 CAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 138  
 XX OY 41 TyrSerGlyProAsnGluArgThrThrPheTrpHisGlySerAsnGluGlyAlaValSer 60  
 XX DB 139 TATAGTGGCCCAATAGAGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 198

OY 61 LeuLeuLeuArgTyrCysAla 67  
 DB 199 CTTTGTACGATATGTGCT 219  
 RESULT 2  
 AAD24022  
 ID AAD24022 standard; CDNA: 1914 BP.  
 XX  
 XX AC AAD24022;  
 XX  
 XX DT 26-MAR-2002 (first entry)  
 XX  
 XX DE Human protein phosphatase-4 CDNA.  
 XX  
 XX KW Human; protein phosphatase; PP-4; immune system disorder; AIDS; allergy;  
 XX KW neurological disorder; developmental disorder; Alzheimer's disease;  
 XX KW cell proliferative disorder; Huntington's disease; arteriosclerosis;  
 XX KW renal tubular acidosis; gonadal dysgenesis; cancer; adenocarcinoma;  
 XX KW Leukemia; transgenic animal; gene therapy; ss.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX FH Key Location/Qualifiers  
 XX FT 203..1465  
 XX FT CDS /tag- a  
 XX FT /product- "Protein phosphatase-4"  
 XX  
 XX PN MO200196546-A2.  
 XX  
 XX PD 20-DEC-2001.  
 XX  
 XX PF 14-JUN-2001; 2001WO-US19442.  
 XX  
 XX PR 16-JUN-2000; 2000US-212447P.  
 XX PR 22-JUN-2000; 2000US-213746P.  
 XX PR 29-JUN-2000; 2000US-215210P.  
 XX PR 06-JUL-2000; 2000US-216529P.  
 XX PR 12-JUL-2000; 2000US-218080P.  
 XX PR 21-JUL-2000; 2000US-220117P.  
 XX  
 XX PA (INCYTE) INCYTE GENOMICS INC.  
 XX  
 XX PI Au-Young J, Baughn MR, Ding L, Elliott VS, Gandhi AR, Griffin JN;  
 XX PI Hafalia A, Kearney L, Lee EA, Lu Y, Nguyen DB, Patterson C;  
 XX PI Rankmar J, Reddy R, Sanjanwala MS, Stewart EA, Tang YI;  
 XX PI Thornton M, Tribouley CM, Walla NK, Yang J, Yao MG, Yue H;  
 XX  
 XX DR MPI: 2002-090206/12.  
 XX DR P-PSDB; AAE14454.  
 XX  
 XX PT Novel polypeptide, useful for diagnosing, treating or preventing  
 XX PT disorders of growth and development, immune system, neurological and  
 XX PT cell proliferation diseases, comprises cancer protein phosphatase  
 XX PT polypeptides  
 XX  
 XX Claim 5; Page 112-113; 116pp; English.  
 XX  
 XX The present sequence is human protein phosphatase (PP)-4 CDNA.  
 XX PP polynucleotide and polypeptide are useful in the diagnosis,  
 XX treatment and prevention of immune system disorders, neurological  
 XX disorders, developmental disorders and cell proliferative disorders.  
 XX Examples of immune system disorders include acquired immune deficiency  
 XX syndrome (AIDS), severe combined immunodeficiency disease (SCID),  
 XX adult respiratory distress syndrome, allergies, amyloidosis,  
 XX anaemia, asthma, atherosclerosis, Crohn's disease, atopic  
 XX dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
 XX Graves' disease, multiple sclerosis, myasthenia gravis, myocardial or  
 XX pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,  
 XX psoriasis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome,  
 XX scleroderma, systemic sclerosis, trauma; neurological disorders include  
 XX Alzheimer's disease, Huntington's disease, dementia, epilepsy,  
 XX Parkinson's disease, mental retardation and other developmental  
 XX disorders of central nervous system such as Down's syndrome, cerebral

CC palsy, periodic paralysis, mental disorders including mood, anxiety,  
 CC and schizophrenic disorders, seasonal affective disorder such as  
 CC akathisia, amnesia, cataplexy, dyskinesia; developmental disorders  
 CC include e.g. renal tubular acidosis, Duchenne and Becker muscular  
 CC dystrophy, gonadal dysgenesis, hypothyroidism; cell proliferative  
 CC disorders include e.g. actinic keratosis, arteriosclerosis,  
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis and  
 CC cancer including adenocarcinoma, leukemia. The polypeptide and  
 CC polynucleotide are further useful for analysing proteome of a tissue  
 CC or a cell type, for screening an agonist/antagonist, a compound that  
 CC specifically binds to it or its modulator. The polynucleotide is useful  
 CC for creating knockin humanised animals (pigs) or transgenic animals  
 CC (mice or rats) to model human disease, for generating a transcript image  
 CC of a tissue or cell type, which represents the global pattern of gene  
 CC expression by a particular tissue or cell type.

CC Sequence 1914 BP; 565 A; 378 C; 438 G; 533 T; 0 other;

# Alignment Scores:

Pred. No.:	2,596-23	Length:	1914
Score:	239.50	Matches:	66
Percent Similarity:	26.29%	Conservative:	0
Best Local Similarity:	26.29%	Mismatches:	1
Query Match:	64.91%	Indels:	185
DB:	24	Gaps:	1

US-09-935-703-7 (1-67) x AAD24022 (1-1914)

QY 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyLupheAsnSergLysn 20  
 DB ATGCAAGATTTATGGCTTACACTAGAAATCTCCGTGTAGTTCACTCTGGAAAT 739  
 QY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuPro----- 35  
 DB 740 CAACCAACCAACAGAGAAAAAACAAGATACCGAGATATCTTCATATGATCAACACGC 799  
 QY 35 ----- 35  
 DB 800 GTTCTCTTGGAAAAAGCAAGACATACATCATGCTAGTTATTTAGAAATGATCATTTGT 859  
 QY 35 ----- 35  
 DB 860 GGAGAAGATATTTTATATCTACTACTCAAGACACTGCTGAGCAACATAGATGACTTT 919  
 QY 35 ----- 35  
 DB 920 TGGCAATGGTGTGGAAAAATATTCAAATGTTATGCCATGATTAACAGAGATAGAA 979  
 QY 35 ----- 35  
 DB 980 GGTGAAATTTATCAAAATGCTACCATTAAGTCCCATTTCTCTGTGAAGAACCATTTGGAATTG 1039  
 QY 35 ----- 35  
 DB 1040 AAACACTCCGTATATCTCTGGAGAACTACAGATACCTTCAATATTTTCATTTGCAATG 1099  
 QY 35 ----- 35  
 DB 1100 TTTCAGTTGTGGAGAACTCAACGGGAACTAGTCACTCTGTAAACAGTTGCAAGTTCCACC 1159  
 QY 35 ----- 35  
 DB 1160 AAGTGGCCAGACCATGACCTCTGCTCAGCAGATAGCTTCAATAATATATTCCTTAT 1219  
 QY 35 ----- 35  
 DB 1220 GCAGAGAGAGCCACCTTACAGAGCCATGTTGTCTACTGACAGTCCGCGCATAGGCCGG 1279  
 QY 36 -----Phe 36  
 DB 1280 ACAGGGGTGTTCTTATGTGTGATGTCGTGTTCTGTGCCATCGTAAAGAACTGTTCAATTC 1339  
 QY 37 GlnHisIstGlyTyrSerGlyProAsnGluArgThrThrPheTrpHisIstGlySerAsnGlu 56

DB 1340 AA-CATCTGTGATATAGTGGCCCAATAGAGAACACCTTGTGCAATCAACGAA 1398  
 QY 57 GYAlaValSerLeuLeuLeuArgTyrCysAla 67  
 DB 1399 GGACGAGATCATCTTTTGTTCAGATATTTGTCT 1431

## RESULT 3

AAST0229 standard; cDNA; 3264 BP.

AAST0229;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #6033.

Human: chromosome mapping; gene mapping; gene therapy; forensic;

food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

MO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Dzmanac RT, Liu C, Yang YT;

WPI; 2001-639362/73.

P-PSDB; ABG06042.

Claim 1; SEQ ID No 6033; 103pp; English.

The invention relates to isolated polynucleotide (I) and

polypeptide (II) sequences. (I) is useful as hybridisation probes,

polymerase chain reaction (PCR) primers, oligomers, and for chromosome

and gene mapping, and in recombinant production of (II). The

polynucleotides are also used in diagnostics as expressed sequence tags

for identifying expressed genes. (I) is useful in gene therapy techniques

to restore normal activity of (II) or to treat disease states involving

(II). (II) is useful for generating antibodies against it, detecting or

quantitating a polypeptide in tissue, as molecular weight markers and as

a food supplement. (II) and its binding partners are useful in medical

imaging of sites expressing (II). (I) and (II) are useful for treating

disorders involving aberrant protein expression or biological activity.

The polypeptide and polynucleotide sequences have applications in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits to assess biodiversity

and to produce other types of data and products dependent on DNA and

amino acid sequences. AAS64197-AAS94564 represent novel human

diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed

specification, but was obtained in electronic format directly from WIPO

at ftp://ipo.int/pub/published\_pct\_sequences.

Sequence 3264 BP; 979 A; 604 C; 676 G; 1005 T; 0 other;

Alignment Scores:

Pred. No.:	6,916-22	Length:	3264
Score:	231.50 <td>Matches:</td> <td>65 </td>	Matches:	65
Percent Similarity:	25.90% <td>Conservative:</td> <td>0 </td>	Conservative:	0

Best Local Similarity: 25.908  
Query Match: 62.748  
DB: 23  
Mismatches: 2  
Indels: 185  
Gaps: 1

US-09-935-703-7 (1-67) x AAV81746 (1-3264)

```

OY 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
DB 901 ATTCGAGGATTTATGGCTTGTAGAACTTGTGCGTGGAGTTCTACTCTGGGAAT 960
OY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuPro- 35
DB 961 CAACCAAGCAACAGAGAAAAAAGAGATACCGAGATATTTCTTCATATGATTCACACGCC 1020
OY 35 1021 GTTCTCTTGGAAAAAGCAAGACTACATCATCTAGTATTATATGAAATAGTCAATTGT 1080
DB 35 1081 GGAGAGAGATTTTATATGCTACTCAAGACACTGCTGAGCAGCATAGATGACTTT 1140
OY 35 1141 TGGCAATGCTGTGGAAAAATATTCAAATGTTATGCCATGATTAACAGAGATGGA 1200
DB 35 1201 GGTTGAATTTATCAATGCTACCATTTACTGCGCCATTCTCTGAAGAGCATTTGAATTG 1260
OY 35 1261 AAACACTCCGCTGATTCCTGGAGAACTACAGATCTCAATATTTGATCATTCGAAATG 1320
DB 35 1321 TTTCAGATTGTGAGAGAGTCCAGCGGAAGTACCTCTGTAAACAGTTGACATTCAC 1380
OY 35 1381 AAGTGGCCAGACATGGCAGCTCTGCTCAGCAGATAGCTTCAATAAATATATGCTTAT 1440
DB 35 1441 GCAAGAGAGACCCCTTACAGGACCATGCTTCTTACCTGCAATGCCGAGATAGCCGG 1500
OY 36 1501 ACAGGGGTTCCTATGTTGATGTGATGTGTTCTGTCATCGTAAGGACATTCATTC 1560
DB 36 1561 AA-CATCATGTGATATAGTGGCCCAATGAGAGAACAGTCTGCGCATGTTCAACAGAA 1619
OY 57 GlnAlaValSerLeuLeuLeuArgTyrCysAla 67
DB 1620 GGAGCAGATACACTTTTGTACGATATTGCT 1652

```

RESULT 4  
ID AAV81746 standard; cDNA; 1692 BP.  
AAV81746:  
AC AAV81746:  
DT 10-MAR-1999 (first entry)  
DE Mouse PTP05 isoform #2 encoding cDNA.  
XX PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; ss.  
XX

OS Mus sp.  
XX Key Location/Qualifiers  
FH Key 199.1413  
FT CDS //tag="a  
FT /note="no stop codon given"  
PD 05-NOV-1998.  
PN MO9849317-A2.  
XX  
XX 27-APR-1998; 98MO-US08439.  
XX 23-OCT-1997; 97US-0063595.  
XX 28-APR-1997; 97US-0044428.  
XX 20-MAY-1997; 97US-0047222.  
XX 11-JUN-1997; 97US-0049477.  
XX 11-JUN-1997; 97US-0049756.  
XX 18-JUN-1997; 97US-0049914.  
XX  
XX (SUGEN-) SUGEN INC.  
XX App H, Clary D, Courtneidge SA, Hui TH, Jallat B;  
XX Markoy D, Onrust S, Peles E, Plozman GD;  
XX MPI; 1999-009434/01.  
XX P-PSDB; AAW89251.  
XX  
XX New nucleic acid encoding specific protein tyrosine phosphatases -  
XX useful for identifying specific modulators for treatment and  
XX prevention of cancer and neurodegenerative disease  
XX  
XX Claim 2; Page 148-149; 193pp; English.  
XX  
XX The present invention describes isolated, enriched or purified nucleic  
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
XX present sequence encodes mouse PTP05. The above proteins, other than  
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
XX substances that modulate their activity (i.e. agonists and antagonists,  
XX including NBP) in vivo or in vitro. These substances are used to treat  
XX or prevent diseases associated with abnormal signal transduction  
XX pathways that involve the proteins, particularly cancer (e.g. leukaemia  
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor  
XX serine/threonine kinase) are used to promote neuronal survival.  
XX particularly for treating Alzheimer's, Parkinson's or Huntington's  
XX diseases. Nucleic acid fragments of the polynucleotides encoding the  
XX CC proteins can be used as probes to identify and clone related sequences;  
XX to detect protein-encoded RNA; to generate transgenic animals and in  
XX gene therapy (optionally after mutation). Ab are used to determine the  
XX proteins.  
XX  
XX Sequence 1692 BP; 507 A; 342 C; 386 G; 457 T; 0 other;  
XX  
XX  
XX Alignment Scores:  
XX Pred. No.: 4.21e-07 Length: 1692  
XX Score: 123.00 Matches: 26  
XX Percent Similarity: 66.00% Conservative: 7  
XX Best Local Similarity: 52.00% Mismatches: 16  
XX Query Match: 33.33% Indels: 1  
XX DB: 20 Gaps: 0  
XX  
XX US-09-935-703-7 (1-67) x AAV81746 (1-1692)  
XX  
XX 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20  
DB 694 ATTCGAGATTTTGGAGTTAGAACAAATGACTCTGCTCATGATTCATTTCTGGGAAT 753  
OY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGln-HisHisG1 40  
DB 754 ACACAGACAGAACAGAGATTAAGAACAGATACCGAGATATTTCTTCATATGATTCACACCGT 813  
OY 40 YTYrSerGlyProAsnGluArgThrThr 49  
XX

DB 814 GTTCTCTTGAAAAACAGACTACA 841

RESULT 5  
AAV81744  
ID AAV81744 standard; cDNA; 1785 BP.

XX  
AC AAV81744;  
XX  
DT 10-MAR-1999 (first entry)  
XX  
DE Mouse PTP05 encoding cDNA.  
XX  
KW PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; ss.

OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 199..1479  
ET /tag= a

PN MO9849317-A2.  
XX  
PD 05-NOV-1998.  
XX  
PF 27-APR-1998; 98MO-US08439.  
XX  
PR 23-OCT-1997; 97US-0063595.  
PR 28-APR-1997; 97US-0044428.  
PR 20-MAR-1997; 97US-0047222.  
PR 11-JUN-1997; 97US-0049477.  
PR 11-JUN-1997; 97US-0049756.  
PR 18-JUN-1997; 97US-0049914.

XX  
PA (SUGEN-) SUGEN INC.  
XX  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;  
PI Marbby D, Onrust S, Peles E, Plowman GD;  
XX  
DR WPI: 1999-009434/01.  
XX  
DR P-PSDB: AAW89249.

PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease

XX  
PS Claim 2; Page 147; 193pp; English.

XX  
XX The present invention describes isolated, enriched or purified nucleic  
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
XX present sequence encodes mouse PTP05. The above proteins, other than  
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
XX substances that modulate their activity (i.e. agonists and antagonists,  
XX including NBP) in vivo or in vitro. These substances are used to treat  
XX or prevent diseases associated with abnormal signal transduction  
XX pathways that involve the proteins, particularly cancer (e.g. leukaemia  
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor  
XX serine/threonine kinase) are used to promote neuronal survival.  
XX particularly for treating Alzheimer's, Parkinson's or Huntington's  
XX diseases. Nucleic acid fragments of the polynucleotides encoding the  
XX proteins can be used as probes to identify and clone related sequences;  
XX to detect protein-encoded RNA; to generate transgenic animals and in  
XX gene therapy (optionally after mutation). Ab are used to determine the  
XX proteins.

XX  
SQ Sequence 1785 BP; 527 A; 361 C; 408 G; 489 T; 0 other;

Alignment Scores: 4.52e-07 Length: 1785  
Pred. No.: 123.00 Matches: 26  
Score: 66.008 Conservative: 7  
Percent Similarity:

Best Local Similarity: 52.00% Mismatches: 16  
Query Match: 33.33% Indels: 1  
DB: 20 Gaps: 0

US-09-935-703-7 (1-67) x AAV81744 (1-1785)

QY 1 MetGInciuphemelAlaLeucluleuysAsnleuProclYgluPhenSerGlyasn 20  
DB 694 ATCCGTGAGTTTGGAGTTAGAACAAATGACTGTGCTGATGACTCAATTCGGGAAT 753  
QY 21 GlnProSerAsnArgGluLysAsnArgTYrArgAspIleLeuProPhcGln-HisHsc1 40  
DB 754 ACACACAGACAGACAGATAGACAGATACGAGATATCTTCATATGATTCACACACT 813  
QY 40 YTYrSerGlyProAsnGluArgThr 49  
DB 814 GTTCTCTTGAAAAACAGACTACA 841

RESULT 6  
AAV81745  
ID AAV81745 standard; cDNA; 1896 BP.

XX  
AC AAV81745;  
XX  
DE 10-MAR-1999 (first entry)  
XX  
DE Mouse PTP05 isoform #1 encoding cDNA.  
XX  
KW PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; ss.

OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 199..1590  
ET /tag= a

PN MO9849317-A2.  
XX  
PD 05-NOV-1998.  
XX  
PF 27-APR-1998; 98MO-US08439.  
XX  
PR 23-OCT-1997; 97US-0063595.  
PR 28-APR-1997; 97US-0044428.  
PR 20-MAR-1997; 97US-0047222.  
PR 11-JUN-1997; 97US-0049477.  
PR 11-JUN-1997; 97US-0049756.  
PR 18-JUN-1997; 97US-0049914.

XX  
PA (SUGEN-) SUGEN INC.  
XX  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;  
PI Marbby D, Onrust S, Peles E, Plowman GD;  
XX  
DR WPI: 1999-009434/01.  
XX  
DR P-PSDB: AAW89250.

PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease

XX  
PS Claim 2; Page 147-148; 193pp; English.

XX  
XX The present invention describes isolated, enriched or purified nucleic  
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
XX present sequence encodes mouse PTP05. The above proteins, other than  
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
XX substances that modulate their activity (i.e. agonists and antagonists,  
XX including NBP) in vivo or in vitro. These substances are used to treat  
XX or prevent diseases associated with abnormal signal transduction

CC pathways that involve the proteins, particularly cancer (e.g. leukemia  
CC and lymphoma), while modulators of Akt-7 (which is a type I receptor  
CC serine/threonine kinase) are used to promote neuronal survival.  
CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
CC diseases. Nucleic acid fragments of the polynucleotides encoding the  
CC proteins can be used as probes to identify and clone related sequences;  
CC to detect protein-encoded RNA; to generate transgenic animals and in  
CC gene therapy (optionally after mutation). Ab are used to determine the  
CC proteins.

XX Sequence 1896 BP; 572 A; 372 C; 427 G; 525 T; 0 other;

# Alignment Scores:

Pred. No.:	4.9e-07	Length:	1896
Score:	123.00	Matches:	26
Percent Similarity:	66.00%	Conservative:	7
Best Local Similarity:	52.00%	Mismatches:	16
Query Match:	33.33%	Indels:	1
DB:	20	Gaps:	0

US-09-935-703-7 (1-67) x AAV81745 (1-1896)

OY 1 MetGingluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20  
Db 805 ATCCGTGAGTTTGGAGTTAGAACAAATGACTGCGCTGAGGACTCAATCTGGGAAT 864  
OY 21 GlnProSerAsnArgGluLysAsnArgGlyArgAspIleLeuProPheGln-HisHisG 40  
Db 865 ACACACAGAACAGATAGAACAGATACAGATATCTTCATATGATCAACACCGT 924  
OY 40 YTYrSerGlyProAsnGluArgThr 49  
Db 925 GTTCCTCTTGAAAAAACAGGACTACA 952

RESULT 7  
ID ABR84570 standard; cDNA: 2160 BP.

XX ABR84570;

XX 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #1141.

XX Human: ss: granulocytic cell; DNA chip: bacterial infection;  
XX viral infection; parasitic infection; protozoal infection;  
XX fungal infection; sterile inflammatory disease; psoriasis;  
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
XX cardiac reperfusion injury; renal reperfusion injury; ARDS;  
XX adult respiratory distress syndrome; inflammatory bowel disease;  
XX Crohn's disease; ulcerative colitis; periodontal disease;  
XX granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

XX MO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001, 2001MO-US30821.

XX 03-OCT-2000, 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX MPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression  
XX of genes associated with granulocyte activation, which serves as  
XX diagnostic markers that is useful for monitoring disease states and  
XX drug toxicity

XX Claim 1; SEQ ID No 1141, 114pp: English.

XX The invention relates to detecting (M1) granulocyte (GC) activation  
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
XX DNA chip analysis as given in the specification, and comparing  
XX the expression level to an expression level in an unactivated  
XX GC, where differential expression of Gs is indicative of GCA.  
XX Also included are modulating (M2) GA by contacting GC with an agent  
XX that alters the expression of at least one gene in Gs; (2) screening (M3)  
XX for an agent capable of modulating GCA or an inflammation (especially  
XX chronic) in a tissue, an allergic response in a subject, exposure of a  
XX subject to a pathogen or sterile inflammatory disease using the  
XX gene expression profile; (3) detecting (M4) an inflammation (especially  
XX chronic) in a tissue, an allergic response in a subject, exposure of a  
XX subject to a pathogen or sterile inflammatory disease, by detecting the  
XX level of expression in a sample of the tissue of gene(s) from Gs, where  
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
XX an allergic response in a subject, exposure of a subject to a pathogen  
XX or sterile inflammatory disease, by contacting a tissue having  
XX inflammation with an agent that modulates the expression of gene(s)  
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
XX modulating GA; M3 is useful for screening an agent capable of modulating  
XX GCA preferably in an inflammation in a tissue; M4 is useful for  
XX detecting an inflammation (especially chronic) in a tissue, an allergic  
XX response in a subject, exposure of a subject to a pathogen or sterile  
XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
XX reperfusion injury, ARDS, adult respiratory distress syndrome,  
XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
XX periodontal disease; also bacterial infection, viral infection,  
XX parasitic infection, protozoal infection, fungal infection, and M5 is  
XX useful for treating one of the above conditions. The present  
XX sequence represents a gene differentially expressed in granulocytes.  
XX Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic  
XX ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2160 BP; 578 A; 584 C; 561 G; 437 T; 0 other;

# Alignment Scores:

Pred. No.:	0.0387	Length:	2160
Score:	88.50	Matches:	24
Percent Similarity:	37.35%	Conservative:	7
Best Local Similarity:	28.92%	Mismatches:	19
Query Match:	23.98%	Indels:	33
DB:	24	Gaps:	3

US-09-935-703-7 (1-67) x ABR84570 (1-2160)

OY 2 GingluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21

Db 460 GAGGAGTTCACTGATGCGATCTGACACATACAGAACTTTGTAACGGCAATATAA 519

OY 22 ProSerAsnArgGluLysAsnArgGlyArgAspIleLeuProPheGlnHis----- 38

Db 520 GAAGAAACAGAGAAAAAACAGATATCCACATCTCCCAATGACATCTAGAGTG 579

OY 39 -----His 39

Db 580 ATTCTGAGCCAACTGATGGAATTCCTGTTGAGCTACATCAATGCTTCTACATAGAT 639

OY 40 GlyTYr-----SerGlyProAsnGluArgThr----- 49

Db 640 GGTACAAAGACAGAAATTAATTCATAGCAGCTCAAGCTCCCAACAGAAACGTTAAC 699

OY 50 ---PheTTP 51

Db 700 GACTTCTGG 708

RESULT 8

AA157987  
ID AA157987 standard; cDNA: 2389 BP.  
AC AA157987;  
XX 22-OCT-2001 (first entry)  
XX  
XX Human polynucleotide SEQ ID NO 190.  
DE  
XX Human; neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia; ss.  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.  
XX 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao QA, Zhou F, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
XX P-PSDB: AAM38831.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
XX  
XX Claim 1; SEQ ID NO 190; 10078bp; English.  
XX  
XX The invention relates to human nucleic acids (AA15798-A161369) and  
XX the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localized neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilization of the activities such as: immune system suppression,  
XX CC Activin/Inhibin activity, chemotactic/chemokine activity, haemostatic  
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
XX CC assays for receptor activity, arthritis and inflammation, leukaemia and  
XX C.N.S disorders.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification.  
XX  
XX Sequence 2389 BP; 656 A; 612 C; 584 G; 537 T; 0 other;  
XX  
XX Alignment Scores:  
XX Pred. No.: 0.0443 Length: 2389  
XX Score: 88.50 Matches: 24  
XX Percent Similarity: 37.35% Conservative: 7  
XX Best Local Similarity: 28.92% Mismatches: 19  
XX Query Match: 23.98% Indels: 33  
XX DB: 22 Gaps: 3

US-09-935-703-7 (1-67) x AA157987 (1-2389)  
QY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21  
Db 460 GAGGAGTCAACATCATGTCATCTGCACACATACAGGACACTTGTGAACGTGGCAAAATAA 519  
QY 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis----- 38  
Db 520 GAAGAAACAGAGAAAAAACAAGATATCCCAATCCTTCCCAATGACCAATTCAGGCTG 579  
QY 39 -----His 39  
Db 580 ATTCTGAGCCACATCGATGATGCAATTCCTGTCAGACATCATCAATGCTTCTACATAGAT 639  
QY 40 GlyTyr-----SerGlyProAsnGluArgThr---- 49  
Db 640 GGTACAAAGAGAAATTAATTCATAGCAGCTCAGGTCCCAAGAAAGGTTAAC 699  
QY 50 ---PheTyr 51  
Db 700 GACTCTGG 708  
RESULT 9  
ID AA159773 standard; cDNA: 2393 BP.  
XX  
XX AA159773;  
XX  
XX 22-OCT-2001 (first entry)  
XX  
XX Human polynucleotide SEQ ID NO 3762.  
XX  
XX  
XX Human; neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.  
XX 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao QA, Zhou F, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
XX P-PSDB: AAM40617.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
XX  
XX Claim 1; SEQ ID NO 3762; 10078bp; English.  
XX  
XX The invention relates to human nucleic acids (AA15798-A161369) and  
XX the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,

CC	Immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic CC and thrombolytic activity, cancer diagnosis and therapy, drug screening, CC assays for receptor activity, arthritis and inflammation, leukaemias and CC C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed CC specification.
XX	
SO	Sequence 2393 BP; 634 A; 655 C; 615 G; 489 T; 0 other;
	Alignment Scores:
	Pred. No.: 0.0444 Length: 2393
	Score: 88.50 Matches: 24
	Percent Similarity: 37.35% Conservative: 7
	Best Local Similarity: 28.92% Mismatches: 19
	Query Match: 23.98% Indels: 33
	Gaps: 3
US-09-935-703-7 (1-67) x AA159773 (1-2393)	
QY	2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21
Db	460 GAGGAGTTCACACTCTATTCGCATCTCGACACATACAGCAACTTTCACTGGCAATATA 519
QY	22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis----- 38
Db	520 GAAGAAAACAGAGAAAAAAGAGATATCCACACATCTCCCAATGACCATCTAGGATG 579
QY	39 -----His 39
Db	580 ATTCTGAGCCAACTGGATGAAATTCCTGTTCAGACTACATCAATGCTTCTACATATGAT 639
QY	40 GlyTyr-----SerGlyProAsnGluArgThr----- 49
Db	640 GGTACCAAGAGAGAAATTAATTCATACAGCTCAAGAGTCCCAACAGAAACGGTTAAC 699
QY	50 ---PheTrp 51
Db	700 GACTTCTGTG 708
RESULT 10	
ID	AAH98486
XX	AAH98486 standard; cDNA; 2689 BP.
XX	
AC	AAH98486;
XX	
DT	12-OCT-2001 (first entry)
XX	
DE	Human EST-derived coding sequence SEQ ID NO: 343.
XX	
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition; ss.
XX	
OS	Homo sapiens.
XX	
PN	MO200154477-A2.
XX	
PD	02-AUG-2001.
XX	
PF	25-JAN-2001; 2001MO-US02687.
XX	
PR	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
XX	

```

PR 15-SEP-2000; 200005-0663870.
xx (HYSE-) HYSEQ INC.
xx
xx Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
xx PI Cao Y, Drmanac RA, Zhang J, Werhman T;
xx DR MPI: 2001-476164/51.
xx P-PSDB; AAM23827.
xx
xx Isolated polypeptide for treatment of diseases, diagnostics, raising
xx PT antibodies and research use -
xx PS
xx Claim 1; Page 432; 1275pp; English.
xx
xx The present invention provides the protein and coding sequences of novel
xx CC proteins from a variety of organisms, including human, dog, cat, horse,
xx CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
xx CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
xx CC from the organism of interest. They can be used in diagnostics,
xx CC forensics, gene mapping, identification of mutations, to assess
xx CC biodiversity and for nutritional purposes. The present sequence is a cDNA
xx CC of the invention.
xx
SQ Sequence 2689 BP; 713 A; 706 C; 685 G; 585 T; 0 other;

Alignment Scores:
Pred. No.: 0.0518 Length: 2689
Score: 88.50 Matches: 24
Percent Similarity: 37.35% Conservative: 7
Best Local Similarity: 28.92% Mismatches: 19
Query Match: 23.98% Indels: 33
DB: 22 Gaps: 3

US-09-935-703-7 (1-67) x AAH98486 (1-2689)

QY 2 GINGLUPHEMELALEULCULULYSANLEUPROGLYLUPHEASNSERGLYASNGLN 21
Db ::::|||||:||||| :::: ||| ||| |||:
756 GAGGGATCTCACTCATCTGGCATCTGGACACATCAAGAACTTTTGACTGGCAATTA 815
QY 22 PROSERASNAATGGLULYSANATGTYTATAGASPILEUPROPHGLNHS----- 38
Db :|||||:|||||:|||||:|||||:|||||:|||||
816 GAAGAAAACAGGAAAAAACAAGATATCCCAACATCTTCCCAATGACCATTTCTAGGGTG 875
QY 39 -----His 39
Db 876 ATTCGAGCAACATGATGAATTCCGTTCAGACTACATCATGCTCTACATAGAT 935
QY 40 GLIYYT-----SERGLYPROASNGLUATGTHRTHR--- 49
Db :|||||:|||||:|||||:|||||:|||||:|||||
936 GGTTCACAGAGAGAAATATTAATTCATAGACGCTCAAGGTCCCAACAGGAAACGGTTAC 995
QY 50 ---PheTrp 51
Db :|||||
996 GACTTCTGG 1004

RESULT 11
AB199311
ID AB199311 standard; cDNA: 5425 BP.
AC
AB199311;
XX
XX 07-MAR-2002 (first entry)
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:171.
XX
XX Mouse; Ischaemia; compressive ischaemia; occlusive ischaemia;
KM vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
XX Mus musculus.
XX
XX WO200188188-A2.
XX
XX
```





Db 491 GAGGAGTTGAGAGTGTGCAAAAAGACAGAGTAAGAACTGCACACGCTCTG---GAA 547  
 Oy 19 GtYAsnGlnProSerAsnArgTylusAsnArgTyrArgAspIleuProPheGlnHis 38  
 548 GGGCAGCGCGCGGAGACAGACAGACCGCTACAGAACATCTTCCCTTGACAC 607

## RESULT 13

AAT35310  
 ID AAT35310 standard; cDNA; 1624 BP.

XX AAT35310;

XX 13-NOV-1996 (first entry)

DE Human SH-PTP1 cDNA clone derived from erythroleukemia cell line.

XX PTP; protein tyrosine phosphatase; SH2; Src homology region 2;

XX chromosome 12p; abnormality; mutation; detection; probe; neoplasia;

XX cancer; leukaemia; diagnosis; megakaryocyte regulation; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 145..2040

XX US5536636-A.

XX 16-JUL-1996.

XX 26-JUN-1991; 91US-0721112.

XX 28-FEB-1994; 94US-0202389.

XX 26-JUN-1991; 91US-0721112.

XX 31-JAN-1992; 92US-0829141.

XX 01-DEC-1992; 92US-0983926.

XX (BETH-) BETH ISRAEL HOSPITAL ASSOC.

XX (MAST) MASSACHUSETTS INST TECHNOLOGY.

XX Freeman RM, Neel BG, Plutzky J, Rosenberg RD;

XX WPI; 1996-341506/34.

XX P-PSDB; AAR99316.

XX Detecting 12p chromosomal abnormality associated with neoplastic

XX disease - using SH-PTP1 protein tyrosine phosphatase gene specific

XX probe

XX Example 1; Column 43-48; 63pp; English.

XX AAT35310 is a human cDNA sequence encoding a variant SH-PTP1 (protein

XX tyrosine phosphatase-1, with two SH2 domains) obtained from a human

XX erythroleukemia cell line. The variant sequence lacks a sequence

XX of two nucleotides. AG, corresponding to positions 1868 and 1869 of

XX the wild-type sequence (see AAT35306). A fragment complementary to the

XX chromosome 12p13 abnormality associated with neoplastic disease, in

XX paric, acute lymphoblastic leukaemia. The probe hybridises to a part

XX of the region coding for the two tandem SH2 domains (see AAR99112).

XX If the probe will not hybridise DNA of chromosome 12p13 from a patient

XX sample it is indicative of an abnormality, normally associated with

XX neoplasia. Alternatively the wild-type SH-PTP1 or SH-PTP2 gene or

XX protein may be used for comparison to sequenced PTP genes taken from a

XX patient, where differences indicate an abnormality. The activity of

XX SH-PTP1 may also be purposefully altered by mutation to effect a change

XX in megakaryocyte function and hence platelet production.

XX Sequence 1624 BP; 381 A; 463 C; 486 G; 294 T; 0 other;

XX Alignment Scores: 0.0592 Length: 1624

Score: 86.00 Matches: 19  
 Percent Similarity: 65.00% Conservative: 7  
 Best Local Similarity: 47.50% Mismatches: 10  
 Query Match: 23.31% Indels: 4  
 Gaps: 2

US-09-935-703-7 (1-67) x AAT35310 (1-1624)

Oy 2 GtngluphemeAlaLeu-----GtuleuLysAsnLeuProGlyGluPheAsnSer 18

Db 880 GAGGAGTTGAGAGTGTGCAAGACAGAGAGTGAAGACTTGACACGCTCTG---GAA 936

Oy 19 GtYAsnGlnProSerAsnArgTylusAsnArgTyrArgAspIleuProPheGlnHis 38

Db 937 GGGCAGCGCGCGGAGACAGACAGACCGCTACAGAACATCTTCCCTTGACAC 996

## RESULT 14

AAT26454  
 ID AAT26454 standard; cDNA; 1788 BP.

XX AAT26454;

XX 26-MAR-2001 (first entry)

XX SHP-1 DNA.

XX SHP-2; SHP-1; Src Homology-2; protein tyrosine phosphatase; mutant;

XX neoplastic disorder; obesity; angiogenesis; cancer; immune;

XX hematopoietic; allergy; ss.

XX Undenitified.

XX US6156551-A.

XX 05-DEC-2000.

XX 05-JUN-1998; 98US-0092443.

XX 05-JUN-1998; 98US-0092443.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX (JOSL-) JOSLIN DIABETES CENT.

XX Neel BG, Shoelson S, Pluskey S, O'Reilly AM;

XX WPI; 2001-060166/07.

XX Mutant SH2 domain-containing protein tyrosine phosphatase, useful in in

XX vitro assays to screen for binding partners, inhibitors of tyrosine

XX phosphatase and for treating tyrosine phosphatase-mediated diseases

XX Disclosure; Fig 6; 161pp; English.

XX The present invention relates to an activated SH2 (Src Homology-2)

XX domain containing protein tyrosine phosphatase (SHP-1 or SHP-2)

XX mutant with a mutation in the SH2-domain. Activated mutants of

XX SH2-domain-containing protein tyrosine phosphatases are useful in

XX in vitro assays to screen for binding partners and inhibitors of

XX the phosphatase and in the treatment of PTP-mediated diseases or

XX conditions in a mammal, including neoplastic disorders, obesity and

XX CC to inhibit angiogenesis. Inhibitors identified using the activated

XX mutants are useful for the treatment of cancer, immunosuppression,

XX immunostimulation, hematopoietic stimulation and anti-allergy

XX treatment.

XX Sequence 1788 BP; 437 A; 490 C; 550 G; 311 T; 0 other;

XX Alignment Scores: 0.0673 Length: 1788

XX Pred. No.: 86.00 Matches: 19

XX Score: 86.00 Conservative: 7

XX Percent Similarity: 65.00% Mismatches: 10

XX Best Local Similarity: 47.50% Indels: 4

XX Query Match: 23.31% Gaps: 2

DB:	22	Gaps:	2
US-09-935-703-7 (1-67) x AAF26454 (1-1788)			
QY	2	GingluphemeLaleu-----GluteLysAsnLeuproGlygIuphenaSer	18
	:::	:::	
Db	736	GAGGAGTTTGAGACGTTTCGAGAAACGAGGAGGTGAAACCTTGACACACGCGCTCG--GAA	792
QY	19	GlyAsngInProSerAsnArgIuLysAsnArgTYrArgAspIleLeupropheGInHis	38
		:::	
Db	793	GGGCAACGGCCAGACACACAAAGGCGCAAGAACCCCTACAAACACTTCTCCCTTTGACAC	852

```

0y      2  GlnGluPheMetAlaLeu-----GluLeuTyrAsnLeuProGlyGluPheAsnSer 18
      :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      736  GAGGAGTGTGAGAGATTTCGAGAAGCAGGAGGCTAACAACATTGCACCAAGCTCTG---GAA 793
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
0y      19  GlnAangInProSerAspArgGluTyrAsnArgTyrArgAspIleLeuProPheGlnHis 38
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      793  GGGCAACGGCCAGAGACAAAGGGCAAGAACCGCTAACAAAGACATCTCCCTTTGACCAC 852
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Search completed: June 11, 2003, 12:55:22
Job time : 215 secs

```

Search completed: June 11, 2003, 12:55:22  
Job time : 215 secs

RESULT 15  
 ID AAF26466  
 AC AAF26466  
 DT 26-MAR-2001 (first entry)  
 DE SHP-1 activated double compound mutant D59A/D419A DNA.  
 KW SHP-2; SHP-1; Src Homology-2; protein tyrosine phosphatase; mutant;  
 KW neoplastic disorder; obesity; angiogenesis; cancer; immune;  
 KW hematopoietic; allergy; ss.  
 OS Undefined.  
 PN US6156551-A.  
 PD 05-DEC-2000.  
 PF 05-JUN-1998; 98US-0092443.  
 PR 05-JUN-1998; 98US-0092443.  
 RA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 FA (JOSL-) JOSLIN DIABETES CENT.  
 PI Neel BG, Shoelson S, Pluskey S, O'Reilly AM;  
 DR WPI: 2001-060166/07.  
 PT Mutant SH2 domain-containing protein tyrosine phosphatase, useful in  
 PT vitro assays to screen for binding partners, inhibitors of tyrosine  
 PT phosphatase and for treating tyrosine phosphatase-mediated diseases  
 PS  
 PS Disclosure: Fig 18; 161pp; English.  
 CC The present invention relates to an activated SH2 (Src Homology-2)  
 CC domain containing protein tyrosine phosphatase (SHP-1 or SHP-2)  
 CC mutant with a mutation in the SH2-domain. Activated mutants of  
 CC SH2-domain-containing protein tyrosine phosphatases are useful in  
 CC in vitro assays to screen for binding partners and inhibitors of  
 CC the phosphatase and in the treatment of PTP-mediated diseases or  
 CC conditions in a mammal, including neoplastic disorders, obesity and  
 CC to inhibit angiogenesis. Inhibitors identified using the activated  
 CC mutants are useful for the treatment of cancer, immunosuppression,  
 CC immunostimulation, hematopoietic stimulation and anti-allergy  
 CC treatment.  
 CC Sequence 1788 BP; 435 A; 492 C; 550 G; 311 T; 0 other;  
 CC

**Alignment Scores:**

Pred. No.:	0.0673	length:	1788
Score:	86.00	Matches:	19
Percent Similarity:	65.00%	Conservative:	7
Best local Similarity:	47.50%	Mismatches:	10
Query Match:	23.31%	Indels:	4
DB:	22	Gaps:	2

US-09-935-703-7 (1-67) x AAF26466 (1-1788)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 11:26:50 ; Search time 11 Seconds

(Without alignments)  
252.629 Million cell updates/sec

Title: US-09-935-703-7

Sequence: 1 MOEFMALTELKLNLPGEFNSGN.....TTFWHSNEGAVSILRRYCA 67

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.5	24.0	699	1	PTPE_MOUSE
2	88.5	24.0	700	1	PTPE_MOUSE
3	88	23.8	595	1	PTNG_MOUSE
4	86	23.3	595	1	PTNG_MOUSE
5	79.5	21.5	802	1	PTRA_HUMAN
6	78.5	21.3	458	1	PTNI_HUMAN
7	77.5	21.0	796	1	PTRA_MOUSE
8	77	20.9	521	1	PTPI_MOUSE
9	75	20.3	1631	1	PTPI_MOUSE
10	73.5	19.9	1123	1	PTPI_MOUSE
11	73	19.8	845	1	PTPI_MOUSE
12	72	19.5	585	1	PTPI_MOUSE
13	72	19.5	593	1	PTPI_MOUSE
14	72	19.5	593	1	PTPI_MOUSE
15	71	19.2	775	1	PTPI_MOUSE
16	71	19.2	780	1	PTPI_MOUSE
17	71	19.2	802	1	PTPI_MOUSE
18	70	19.0	363	1	PTPI_MOUSE
19	70	19.0	363	1	PTPI_MOUSE
20	70	19.0	415	1	PTPI_MOUSE
21	69	18.7	1152	1	PTPI_MOUSE
22	69	18.7	1175	1	PTPI_MOUSE
23	69	18.7	1175	1	PTPI_MOUSE
24	69	18.7	1304	1	PTPI_MOUSE
25	68.5	18.6	989	1	PTPI_MOUSE
26	68	18.4	432	1	PTPI_MOUSE
27	68	18.4	432	1	PTPI_MOUSE
28	68	18.4	435	1	PTPI_MOUSE
29	68	18.4	1174	1	PTPI_MOUSE
30	67	18.2	434	1	PTPI_MOUSE
31	67	18.2	829	1	PTPI_MOUSE
32	66	17.9	668	1	PTPI_MOUSE
33	63.5	17.2	807	1	PTPI_MOUSE

## ALIGNMENTS

34	63.5	17.2	989	1	PTP3_DICTDI
35	63	17.1	414	1	PTP3_DICTDI
36	63	17.1	593	1	PTNG_MOUSE
37	63	17.1	1189	1	PTNG_MOUSE
38	63	17.1	2029	1	PTNG_MOUSE
39	62	16.8	377	1	PTP2_DICTDI
40	62	16.8	444	1	PTP2_DICTDI
41	62	16.8	1255	1	PTP2_DICTDI
42	62	16.8	2314	1	PTP2_DICTDI
43	62	16.8	2316	1	PTP2_DICTDI
44	61.5	16.7	794	1	PTP2_DICTDI
45	61.5	16.7	1912	1	PTP2_DICTDI

RESULT 1  
PTPE\_MOUSE STANDARD: PRT; 699 AA.  
AC P49446; Q62134; Q62444; Q64496;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-epsilon)  
GN PTPE OR PTPE.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N;  
RX MEDLINE=96064677; PubMed=7592814;  
RA Elson A., Leder P.;  
RT "Protein-tyrosine phosphatase epsilon. An isoform specifically expressed in mouse mammary tumors initiated by v-Ha-ras or neu.";  
RL J. Biol. Chem. 270:26116-26122(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DKA/2;  
RX Mukoyama Y.;  
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; Tissue=Brain, and Lung;  
RA Hou E.W., Li S.L.;  
RT Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 224-332 FROM N.A.  
RC STRAIN=BALB/c; Tissue=Brain;  
RX MEDLINE=93086603; PubMed=1454056;  
RA Schepens J., Zeeuwen P., Wieringa B., Hendriks W.;  
RT "Identification and typing of members of the protein-tyrosine phosphatase gene family expressed in mouse brain.";  
RL Mol. Biol. Rep. 16:241-248(1992).  
RN [5]  
RP SEQUENCE OF 224-332 FROM N.A.  
RC STRAIN=BALB/c; Tissue=Brain;  
RX MEDLINE=95134232; PubMed=7832766;  
RA Hendriks W., Schepens J., Brugman C., Zeeuwen P., Wieringa B.;  
RT "A novel receptor-type protein tyrosine phosphatase with a single catalytic domain is specifically expressed in mouse brain.";  
RL Blochem. J. 305:499-504(1995).  
RN [6]  
RP CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein tyrosine + phosphate.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U35368; AAC5281.1; -  
 CC EMBL; D83484; BAA11927.1; -  
 CC EMBL; U62387; AAB04553.1; -  
 CC EMBL; 223052; CAA80587.1; -  
 CC EMBL; 223053; CAA80588.1; -  
 CC HSSP; P18052; IYFO.  
 CC MGI; MGI:97813; Ptpre.  
 CC InterPro; IPR000387; TYR\_phosphatase.  
 CC InterPro; IPR000242; TYR\_PP.  
 CC Pfam; PF00102; Y-phosphatase. 2.  
 CC PRINTS; PR00700; PTPPHPTASE.  
 CC SMART; SM00194; PTPC; 2.  
 CC PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
 CC PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 2.  
 CC PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 CC Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.  
 KW SIGNAL.  
 FT CHAIN 1 19 POTENTIAL.  
 FT DOMAIN 20 699 PROTEIN-TYROSINE PHOSPHATASE EPSILON.  
 FT TRANSMEM 46 45 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 69 699 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 153 392 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT DOMAIN 393 699 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT ACT\_SITE 334 334 BY SIMILARITY.  
 FT ACT\_SITE 629 629 BY SIMILARITY.  
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 500 500 G -> A (IN REF. 2).  
 FT CONFLICT 506 506 G -> V (IN REF. 2).  
 FT CONFLICT 521 522 IV -> ML (IN REF. 2).  
 FT CONFLICT 606 606 M -> I (IN REF. 1).  
 SQ SEQUENCE 699 AA; 80645 MW; 4D04467438017FEB CRC64;  
 Query Match 24.0%; Score 88.5; DB 1; Length 699;  
 Best Local Similarity 28.9%; Pred. No. 0.0033;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
 QY 2 QEFMALEKLNLPGEFNSGNQPSNREKNRYRDLPPQH-----H 39  
 DB 136 EEFNSLPFGHIOGFELANKNEENREKNRYRDLPPDHCHVILLSQVDGIPCSDYINASYID 195  
 QY 40 GY-----SGPNERTT--FW 51  
 DB 196 GYKEKNKFIAAGPKQETVNDFW 218  
 RESULT 2  
 PTPE\_HUMAN STANDARD; PRT; 700 AA.  
 AC P23469;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-epiilon).  
 GN PTPRE.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=91006018; PubMed=2170109;  
 RA Krueger N.X., Streuli M., Salto H.;  
 RT "Structural diversity and evolution of human receptor-like protein tyrosine phosphatases.";

RL EMBL J. 9:3241-3252(1990).  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O - protein tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X5A134; CAA38069.1; -  
 CC PIR; S12053; S12053.  
 CC HSSP; P18052; IYFO.  
 CC Genew; HGNC:9669; PTPRE.  
 CC MIM; 600926; -  
 CC InterPro; IPR000387; TYR\_phosphatase.  
 CC InterPro; IPR000242; TYR\_PP.  
 CC Pfam; PF00102; Y-phosphatase. 2.  
 CC PRINTS; PR00700; PTPPHPTASE.  
 CC SMART; SM00194; PTPC; 2.  
 CC PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
 CC PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 2.  
 CC PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 CC Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.  
 KW SIGNAL.  
 FT CHAIN 1 19 POTENTIAL.  
 FT DOMAIN 20 700 PROTEIN-TYROSINE PHOSPHATASE EPSILON.  
 FT TRANSMEM 47 46 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 70 700 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 154 393 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT DOMAIN 394 700 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT ACT\_SITE 335 335 BY SIMILARITY.  
 FT ACT\_SITE 630 630 BY SIMILARITY.  
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 700 AA; 80641 MW; D096BCADCEA5708 CRC64;  
 Query Match 24.0%; Score 88.5; DB 1; Length 700;  
 Best Local Similarity 28.9%; Pred. No. 0.0033;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
 QY 2 QEFMALEKLNLPGEFNSGNQPSNREKNRYRDLPPQH-----H 39  
 DB 137 EEFNSLPFGHIOGFELANKNEENREKNRYRDLPPDHSHVILLSQVDGIPCSDYINASYID 196  
 QY 40 GY-----SGPNERTT--FW 51  
 DB 197 GYKEKNKFIAAGPKQETVNDFW 219  
 RESULT 3  
 PTN6\_MOUSE STANDARD; PRT; 595 AA.  
 AC P29351; Q63872; Q63873; Q63874; Q9QVA6; Q9QVA7; Q9QVA8; Q9ROV6;  
 AC Q92163; Q35128;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 6 (EC 3.1.3.48) (Protein-tyrosine phosphatase 1C) (PTP-1C) (hematopoietic cell protein-tyrosine phosphatase) (70Z-SHP) (SH-PTP1).  
 GN PTPN6 OR PTP1C OR HCP OR HCPH.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

CC STRAIN-DBA/2.  
RX MEDLINE-92123209; Pubmed-1732748;  
RA Yi T., Cleveland J.L., Ihle J.N.;  
RT "Protein tyrosine phosphatase containing SH2 domains:  
RT characterization, preferential expression in hematopoietic cells, and  
RT localization to human chromosome 12p12-p13."  
RL Mol. Cell. Biol. 12:836-846(1992).  
RN [12]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-92236615; Pubmed-1373816;  
RA Matthews R.J., Bowne D.B., Flores E., Thomas M.L.;  
RT "Characterization of hematopoietic intracellular protein tyrosine  
RT phosphatases: description of a phosphatase containing an SH2 domain  
RT and another enriched in proline-, glutamic acid-, serine-, and  
RT threonine-rich sequences."  
RL Mol. Cell. Biol. 12:2396-2405(1992).  
RN [13]  
RP SEQUENCE FROM N.A., AND VARIANTS MOTHEATEN AND VIABLE MOTHEATEN.  
RX STRAIN-C57BL/6J; TISSUE=bone marrow;  
CC MEDLINE-93313972; Pubmed-8324828;  
RX Schultz L.D., Schweitzer P.A., Rajan T.V., Yi T., Ihle J.N.,  
RA Matthews R.J., Thomas M.L., Belter D.R.;  
RT "Mutations at the murine motheaten locus are within the hematopoietic  
RT cell protein-tyrosine phosphatase (Hcph) gene."  
RL Cell 73:1445-1454(1993).  
RN [14]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).  
RX MEDLINE-98112780; Pubmed-9445485;  
RA Anselari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M.,  
RA Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W.,  
RA Gibbs R.A.;  
RT "Comparative sequence analysis of a gene-rich cluster at human  
RT chromosome 12p13 and its syntenic region in mouse chromosome 6."  
RL Genome Res. 8:29-40(1998).  
RN [15]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND SUBUNIT.  
RX STRAIN-C3H; TISSUE=adrenal gland;  
CC MEDLINE-93948302; Pubmed-10419485;  
RA Martin A., Tsai H.W., Shulman M.J., Iseman D., Tsai F.W.;  
RT "Murine SHP-1 splice variants with altered Src homology 2 (SH2)  
RT domains. Implications for the SH2-mediated intramolecular regulation  
RT of SHP-1."  
RL J. Biol. Chem. 274:21725-21734(1999).  
RN [16]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RN [17]  
RP SEQUENCE OF 54-68; 128-135; 137-151; 242-252; 278-285; 293-308 AND  
RX 373-382, AND PHOSPHORYLATION.  
RP MEDLINE-93054686; Pubmed-1385421;  
RA Yeung Y.-G., Berg K.L., Pixley F.J., Angeletti R.H., Stanley E.R.;  
RT "Protein tyrosine phosphatase-1C is rapidly phosphorylated in tyrosine  
RT in macrophages in response to colony stimulating factor-1."  
RL J. Biol. Chem. 267:23447-23450(1992).  
CC -1- FUNCTION: PLAYS A KEY ROLE IN HEMATOPOIESIS. THIS PTPASE ACTIVITY  
CC MAY DIRECTLY LINK GROWTH FACTOR RECEPTORS AND OTHER SIGNALING  
CC PROTEINS THROUGH PROTEIN-TYROSINE PHOSPHORYLATION. THE SH2 REGIONS  
CC MAY INTERACT WITH OTHER CELLULAR COMPONENTS TO MODULATE ITS OWN  
CC PHOSPHATASE ACTIVITY AGAINST INTERACTING SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O -> protein  
CC tyrosine + phosphate.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SUBUNIT: Monomer.  
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN HEMATOPOIETIC  
CC CELLS.  
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; are  
CC produced by alternative splicing.  
CC -1- PTM: Phosphorylated on tyrosine residues.  
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.  
CC -----

```
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).
```

---

```
CC -----
DR EMBL; M68902; AAA37796.1; -.
DR EMBL; M90389; AAA40007.1; -.
DR EMBL; S63763; -. NOT_ANNOTATED_CDS.
DR EMBL; S63764; -. NOT_ANNOTATED_CDS.
DR EMBL; S63803; -. NOT_ANNOTATED_CDS.
DR EMBL; AC002397; AAC36009.1; -.
DR EMBL; AC002397; AAC36008.1; -.
DR EMBL; U65955; AAD00152.1; JOINED.
DR EMBL; U65952; AAD00152.1; JOINED.
DR EMBL; U65953; AAD00152.1; JOINED.
DR EMBL; U65954; AAD00152.1; JOINED.
DR EMBL; U65955; AAD00151.1; JOINED.
DR EMBL; U65951; AAD00151.1; JOINED.
DR EMBL; U65952; AAD00151.1; JOINED.
DR EMBL; U65953; AAD00151.1; JOINED.
DR EMBL; U65954; AAD00151.1; JOINED.
DR EMBL; BC012660; AAH12660.1; -.
DR PIR; A44390; A44390.
DR HSSP; P29350; IGWZ.
DR MGD; MG1; G6055; Hcph.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_pp.
DR Pfam; PF00017; SH2; 2.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PTYPHEPTASE.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRODOM; PD000093; SH2; 2.
DR SMART; SM00194; PTPcc; 1.
DR SMART; SM00252; SH2; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
DR PROSITE; PS50001; SH2; 2.
KW Hydrolase; SH2 domain; Repeat; phosphorylation; Alternative splicing.
FT DOMAIN 4 100 SH2 1.
FT FT 110 213 SH2 2.
FT FT 259 514 PROTEIN-TYROSINE PHOSPHATASE.
FT FT ACT_SITE 453 453 BY SIMILARITY.
FT FT VARSPPLIC 1 3 NVR -> MSRG (IN ISOFORM 2).
FT FT VARSPPLIC 1 3 MISSING (IN ISOFORM 3).
FT FT VARIANT 77 99 SVSR -> MSRG (IN ISOFORM 3).
FT FT VARIANT 77 99 EYVTOOQGLDORDGTITLKPT -> VPRPIRWAGGYTA
FT FT VARIANT 100 595 AGCGALD (IN MOTHEATEN (ME)).
FT FT CONFLICT 240 240 MISSING (IN MOTHEATEN (ME)).
FT FT CONFLICT 572 572 K -> R (IN REF. 1 AND 3).
FT FT CONFLICT 586 586 E -> D (IN REF. 6).
SQ SEQUENCE 595 AA; 67559 MW; CFLI7300D03263BD2 CRC64;
```

---

```
Query Match      23.8%; Score 88; DB 1; Length 595;
Best Local Similarity 47.5%; Pred. No. 0.0031;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2
```

---

```
Db          2 QEFMAL--ELKNTLPGEFNSGNQPSNRERKRYRDILPFCH 38
           :||:| | | ||| | | | | | | | | | | | | | | |
AC          246 EEFEESLQKEVKNTLHORL-EGORPENKSKRKRYKNILPFDH 284
```

---

```
RESULT 4
PTN6_HUMAN
ID         PTN6_HUMAN          STANDARD;             PRT;       595 AA.
AC         P29350;
DT         01-DEC-1992 (Rel. 24, Created)
DT         01-DEC-1992 (Rel. 24, Last sequence update)
```

DT 15-JUN-2002 (Ref. 41, last annotation update)  
DE Protein-tyrosine phosphatase, non-receptor type 6 (EC 3.1.3.48)  
DE (Protein-tyrosine phosphatase 1C) (PTP-1C) (Hematopoietic cell  
DE SHP-1).  
GN PTPN6 OR PTP1C OR HCP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OC NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RA MEDLINE-92123209; PubMed-1732748;  
RA Y1 T., Cleveland J.L., Ihle J.N.;  
RT "Protein tyrosine phosphatase containing SH2 domains:  
RT characterization, preferential expression in hematopoietic cells, and  
RT localization to human chromosome 12p12-p13.";  
RL Mol. Cell. Biol. 12:836-846(1992).  
RN [2]  
RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RC TISSUE-Breast;  
RX MEDLINE-91343005; PubMed-1652101;  
RA Shen S.H., Bastien L., Posner B.I., Chretien P.;  
RT "A protein-tyrosine phosphatase with sequence similarity to the SH2  
RT domain of the protein-tyrosine kinases.";  
RL Nature 352:736-739(1991).  
RN [3]  
RP REVISIONS.  
RA Shen S.H., Bastien L., Posner B.I., Chretien P.;  
RL Nature 353:868-868(1991).  
RN [4]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RX MEDLINE-92141214; PubMed-1736296;  
RA Plutsky J., Neel B.G., Rosenberg R.D.;  
RT "Isolation of a src homology 2-containing tyrosine phosphatase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:1123-1127(1992).  
RN [5]  
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).  
RX MEDLINE-95394454; PubMed-765165;  
RA Banville D., Stocco R., Shen S.H.;  
RT "Human protein tyrosine phosphatase 1C (PTPN6) gene structure:  
RT alternate promoter usage and exon skipping generate multiple  
RT transcripts.";  
RL Genomics 27:165-173(1995).  
RN [6]  
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).  
RX MEDLINE-96303695; PubMed-8723724;  
RA Ansari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,  
RT "A gene-rich cluster between the CpA and triosephosphate isomerase  
RT genes at human chromosome 12p13.";  
RL Genome Res. 6:314-326(1996).  
RN [7]  
RP PHOSPHORYLATION.  
RX MEDLINE-95300784; PubMed-7781604;  
RA Li R.Y., Gault F., Ragab-Thomas J.M.F., Chap H.;  
RT "Tyrosine phosphorylation of an SH2-containing protein tyrosine  
RT phosphatase is coupled to platelet thrombin receptor via a pertussis  
RT toxin-sensitive heterotrimeric G-protein.";  
RL EMBO J. 14:2519-2526(1995).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 248-399.  
RX MEDLINE-98447672; PubMed-9774441;  
RA Yang J., Liang X., Niu T., Meng W., Zhao Z., Zhou G.W.;  
RT "Crystal structure of the catalytic domain of protein-tyrosine  
RT phosphatase SHP-1.";  
RL J. Biol. Chem. 273:28199-28207(1998).  
RN [9]  
RP FUNCTION: PLAYS A KEY ROLE IN HEMATOPOIESIS. THIS PTASE ACTIVITY  
CC MAY DIRECTLY LINK GROWTH FACTOR RECEPTORS AND OTHER SIGNALING  
CC PROTEINS THROUGH PROTEIN-TYROSINE PHOSPHORYLATION. THE SH2 REGIONS  
CC MAY INTERACT WITH OTHER CELLULAR COMPONENTS TO MODULATE ITS OWN  
CC PHOSPHATASE ACTIVITY AGAINST INTERACTING SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: protein tyrosine phosphatase + H(2)O -> protein  
CC tyrosine + phosphate.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE)  
CC AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN HEMATOPOIETIC  
CC CELLS.  
CC -1- PTM: PHOSPHORYLATED ON SERINE AND TYROSINE.  
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
CC -----  
DR EMBL: M74903; AAA35963.1; -;  
DR EMBL: X62055; CAA43982.1; -;  
DR EMBL: M72733; AAA36610.1; -;  
DR EMBL: U15528; AAA82880.1; -;  
DR EMBL: U15536; AAA82880.1; JOINED.  
DR EMBL: U15535; AAA82880.1; JOINED.  
DR EMBL: U15534; AAA82880.1; JOINED.  
DR EMBL: U15533; AAA82880.1; JOINED.  
DR EMBL: U15532; AAA82880.1; JOINED.  
DR EMBL: U15531; AAA82880.1; JOINED.  
DR EMBL: U15529; AAA82880.1; JOINED.  
DR EMBL: U15528; AAA82879.1; JOINED.  
DR EMBL: U15537; AAA82879.1; JOINED.  
DR EMBL: U15535; AAA82879.1; JOINED.  
DR EMBL: U15534; AAA82879.1; JOINED.  
DR EMBL: U15533; AAA82879.1; JOINED.  
DR EMBL: U15532; AAA82879.1; JOINED.  
DR EMBL: U15531; AAA82879.1; JOINED.  
DR EMBL: U15530; AAA82879.1; JOINED.  
DR EMBL: U15529; AAA82879.1; JOINED.  
DR EMBL: U47924; AAB51323.1; -;  
DR EMBL: U47924; AAB51322.1; -;  
DR PIR: S20825; S20825.  
DR PIR: BA2031; BA2031.  
DR PDB: 1GM2; 22-AUG-99.  
DR PDB: 1EPR; 07-MAR-01.  
DR Genew: HGNC:9658; PTPN6.  
DR MIM: 176883; -;  
DR InterPro: IPR000980; SH2.  
DR InterPro: IPR000387; TYR\_phosphatase.  
DR InterPro: IPR000242; TYR\_PP.  
DR Pfam: PF00017; SH2; 2.  
DR Pfam: PF00102; Y\_phosphatase; 1.  
DR PRINTS: PR00700; PTPPHPTASE.  
DR PRINTS: PR00401; SH2DOMAIN.  
DR PRODOM: PD000093; SH2; 2.  
DR SMART: SM00194; PTPc; 1.  
DR SMART: SM00252; SH2; 2.  
DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 1.  
DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
DR PROSITE: PS50001; SH2; 2.  
KW Hydroxylase; SH2 domain; Repeat; Phosphorylation; Alternative splicing;  
KW 3D-structure.  
FT DOMAIN 4 100 SH2 1.  
FT DOMAIN 110 213 SH2 2.  
FT ACT\_SITE 269 514 PROTEIN-TYROSINE PHOSPHATASE.  
FT ACT\_SITE 453 453  
FT VARSPIC 1 39  
FT VARSPIC 40 44  
FT CONFLICT 1 3  
FT CONFLICT 6 6  
H -> L (IN REF. 5).  
MISSING (IN SHORT ISOFORM).  
SLAVR -> MLSRG (IN SHORT ISOFORM).  
MVR -> L (IN REF. 2).  
H -> L (IN REF. 5).

```

FT FT CONFLICT 86 86 L -> V (IN REF. 4).
FT CONFLICT 146 146 V -> E (IN REF. 5).
SO SEQUENCE 595 AA: 67561 MW: 407736C21D3542D2 CRC64:

Query Match 23.3%, Score 86; DB 1; Length 595;
Best Local Similarity 47.5%; Pred. No. 0.0055;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2

OY 2 OEPMAL---ELKKNLGEFNGSCNOPSREKKNYRDLIPQCH 38
DB 246 EEFEESLQKEVKNLHQR-EGORPENKGNKRNKILPIDH 284

RESULT 5
ID PTAA HUMAN STANDARD: PRT: 802 AA.
AC P18433: 014513:
DT 01-NOV-1980 (Rel. 16, Created)
DT 01-NOV-1980 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase alpha precursor (EC 3.1.3.48) (R-PMP-
DE alpha).
GN PTPA OR PTPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90349565: PubMed-2166945;
RA Sap J., D'Eustachio P., Glyol D., Schlesinger J.:
RT "Cloning and expression of a widely expressed receptor tyrosine
RT phosphatase."
RL Proc. Natl. Acad. Sci. U.S.A. 87:6112-6116(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-90384936: PubMed-2169617;
RA Kaplan R., Morse B., Huebner K., Croce C., Howk R., Ravera M.,
RA Rlica G., Jaye M., Schlesinger J.:
RT "Cloning of three human tyrosine phosphatases reveals a multigene
RT family of receptor-linked protein-tyrosine-phosphatases expressed in
RT brain."
RL Proc. Natl. Acad. Sci. U.S.A. 87:7000-7004(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-91006018: PubMed-2170109;
RA Krueger N.X., Streuli M., Saito H.:
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases."
RL EMBO J. 9:3241-3252(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-91088320: PubMed-2175890;
RA Ohagi S., Nishi M., Steiner D.F.:
RT "Sequence of a cDNA encoding human LMP (leukocyte common antigen-
RT related peptide)."
RL Nucleic Acids Res. 18:7159-7159(1990).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE-91032191: PubMed-2172030;
RA Jirik F.R., Janzen N.M., Melhado I.G., Harder K.W.:
RT "Cloning and chromosomal assignment of a widely expressed human
RT receptor-like protein-tyrosine phosphatase."
RL PNAS Lett. 273:239-242(1990).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE-21638749: PubMed-11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Besley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,

```

FT CONFLICT 122 122 E -> P (IN REF. 5).  
 FT CONFLICT 138 138 S -> SCNSKSKDR (IN REF. 2).  
 FT CONFLICT 179 187 MISSING (IN REF. 2, 3, 4, 5 AND 6).  
 FT CONFLICT 289 289 G -> E (IN REF. 4).  
 FT CONFLICT 367 367 V -> A (IN REF. 4).  
 FT CONFLICT 453 453 F -> S (IN REF. 4).  
 FT CONFLICT 786 786 K -> E (IN REF. 4).  
 SQ SEQUENCE 802 AA; 90599 MW; 8E964CB565BE32 CRC64;

Query Match  
 Best Local Similarity 21.5%; Score 79.5; DB 1; Length 802;  
 Matches 22; Conservative 9; Mismatches 19; Indels 33; Gaps 3;

OY 2 QEFMALELNKLPGEFNSGNOPSREKRRYDILPFQH-----H 39  
 DB 243 EEFNLPACPIQATECAASKERKRRYVILPYDHSRVHLPEGVPSDDYINASFIN 302  
 OY 40 GY-----SGPNERTT--FM 51  
 DB 303 GYQEKNFIAAGPKKEFTVDFW 325

## RESULT 6

PTNL\_HUMAN  
 ID PTNL\_HUMAN STANDARD; PRT; 458 AA.  
 AC 099952;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 18 (EC 3.1.3.48)  
 GN PTPN18 OR BDP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCBL\_TaxID-9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND CHARACTERIZATION.  
 RC TISSUE=Brain;  
 RA MEDLINE=97108674; PubMed=8950995;  
 RA Kim Y.W., Wang H.Y., Sures I., Lammer R., Martell K.J., Ullrich A.;  
 RT "Characterization of the PEST family protein tyrosine phosphatase  
 BDP1.";  
 RL Oncogene 13:2275-2279(1996).  
 CC -1- FUNCTION: Differentially dephosphorylate autophosphorylated  
 CC tyrosine kinases which are known to be overexpressed in tumor  
 CC tissues.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- TISSUE SPECIFICITY: Expressed in brain, colon and several tumor-  
 CC derived cell lines.  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; X79568; CAA56105.1; -;  
 DR HSSP; 006124; 2SHP.  
 DR Genew; HGNC:9649; PTPN18.  
 DR MIM; 606587;  
 DR InterPro; IPR000387; Tyr\_phosphatase.  
 DR InterPro; IPR000242; Tyr\_PP.  
 DR Pfam; PF00102; Y\_phosphatase\_1.  
 DR PRINTS; PRO0700; PRTYPHPRASE.  
 DR SMART; SM00194; PTPC; 1.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.

DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 KM Hydrolase.  
 FT DOMAIN 26 291 PROTEIN-TYROSINE PHOSPHATASE.  
 FT ACT\_SITE 229 229 BY SIMILARITY.  
 SQ SEQUENCE 458 AA; 50384 MW; 46BCA1E17C2C78B1 CRC64;

Query Match  
 Best Local Similarity 21.3%; Score 78.5; DB 1; Length 458;  
 Matches 26; Conservative 10; Mismatches 19; Indels 51; Gaps 4;

OY 12 LPGEF-----SGNOPSREKRRYDILPF-----OHGYS- 42  
 DB 26 LAGEFSDIQACSAAMKADVCSTVAGSRPEYKRRKRYDLPDQTRVILSLQEGHSD 85  
 OY 43 -----GPNERT--TFHGSNEGAVSLLRIC 66  
 DB 86 YINGNFIRGVDSLAYIATGCPPLRTLDLFWRLWEGFVKVILMAC 131

## RESULT 7

PTRA\_RAT  
 ID PTRA\_RAT STANDARD; PRT; 796 AA.  
 AC 003348;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase alpha precursor (EC 3.1.3.48) (R-PTP-  
 DE alpha).  
 GN PTPRA OR LRP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCBL\_TaxID-10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA MEDLINE=93038682; PubMed=1417854;  
 RA Moriyma T., Fujiwara Y., Imai E., Takenaka M., Kawanishi S.,  
 RA Inoue T., Noguchi T., Tanaka T., Kamada T., Ueda N.;  
 RT "cDNA cloning of rat LRP, a receptor like protein tyrosine  
 RT phosphatase, and evidence for its gene regulation in cultured rat  
 RT mesangial cells.";  
 RL Biochem. Biophys. Res. Commun. 188:34-39(1992).  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; L01702; AAA41983.1; -;  
 DR HSSP; P18052; IYFO.  
 DR InterPro; IPR000387; Tyr\_phosphatase.  
 DR InterPro; IPR000242; Tyr\_PP.  
 DR Pfam; PF00102; Y\_phosphatase\_2.  
 DR PRINTS; PRO0700; PRTYPHPRASE.  
 DR SMART; SM00194; PTPC; 2.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 KW Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat;  
 KW signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 796 PROTEIN-TYROSINE PHOSPHATASE ALPHA.  
 FT DOMAIN 20 145 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 146 169 POTENTIAL.  
 FT DOMAIN 170 796 CYTOPLASMIC (POTENTIAL).



FT DOMAIN 234 494 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT ACT\_SITE 436 796 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT ACT\_SITE 436 796 BY SIMILARITY.  
 FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 796 AA; 90260 MW; 4793796191056920 CRC64;  
 Query Match 21.0%; Score 77.5; DB 1; Length 796;  
 Best Local Similarity 25.3%; Pred. No. 0.088;  
 Matches 21; Conservative 10; Mismatches 19; Indels 33; Gaps 3;  
 QY 2 QEFMALELNLPGEFNSGNQPSNREKRRRDILPFOH-----H 39  
 DB 237 EEFNLAPACPIATCEAAKSEKREKRYVLLPYDHSRVHLTPVEGVPSDYINASFIN 296  
 QY 40 GT-----SGPNERT--FW 51  
 DB 297 GYQENKRFIAOGPKETVDFW 319  
 RESULT 8  
 PP1L\_DICDI STANDARD; PRT; 521 AA.  
 ID PP1L\_DICDI  
 AC P34137;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Protein-tyrosine phosphatase 1 (EC 3.1.3.48) (Protein-tyrosine-phosphatase phosphodiesterase 1).  
 GN PTPI OR PTPI.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellidae; Dictyostellium.  
 OX NCBI\_Taxid=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-93046662; PubMed-1423620;  
 RA Howard P.K., Sefton B.M., Firtel R.A.;  
 RT Analysis of a spatially regulated phosphotyrosine phosphatase identifies tyrosine phosphorylation as a key regulatory pathway in Dictyostelium. \*;  
 RT Cell 71:637-647(1992).  
 RL -1- FUNCTION: MAY HAVE A ROLE IN GROWTH AND IN THE EARLY STAGES OF DEVELOPMENT. AFFECTS THE TIMING OF DEVELOPMENT.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MIGHT BE TETHERED TO THE PLASMA MEMBRANE.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN ANTERIOR-LIKE CELLS AND TO A LESSER DEGREE IN PRESTALK CELLS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT A VERY LOW LEVEL IN VEGETATIVE CELLS, INDUCED BY 4 HRS, MAXIMALLY EXPRESSED AT THE TIGHT AGGREGATE STAGE AND THROUGH THE REMAINDER OF DEVELOPMENT.  
 CC -1- MISCELLANEOUS: THE PTASE DOMAIN IS INTERRUPTED BY A PTASE INSERT WHICH SHARES NO HOMOLOGIES WITH OTHER PTASE PROTEINS.  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-TYROSINE PHOSPHATASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L07125; AAA33241.1; .  
 DR HSSP; Q06124; 25HP.  
 DR Dictydb; DD03018; ptpa.  
 DR InterPro; IPR000340; DS\_phosphatase.  
 DR InterPro; IPR000387; TYR\_phosphatase.  
 DR InterPro; IPR000242; TYR\_PP.  
 DR Pfam; PF00102; Y\_phosphatase; 1.  
 DR PRINTS; PR00700; PRTYPPHRTASE.  
 DR SMART; SM00194; PTPc; 1.  
 DR SMART; SM00012; PTPc-DSPc; 1.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR Hydrolase.  
 FT DOMAIN 1 114 SER-RICH  
 FT ACT\_SITE 310 310 BY SIMILARITY.  
 FT DOMAIN 327 425 PTASE INSERT (ASN-RICH).  
 FT DOMAIN 382 400 POLY-ASN.  
 SQ SEQUENCE 521 AA; 59427 MW; 0F516AEDD75EAB96 CRC64;  
 Query Match 20.9%; Score 77; DB 1; Length 521;  
 Best Local Similarity 43.2%; Pred. No. 0.062;  
 Matches 16; Conservative 5; Mismatches 16; Indels 0; Gaps 0;  
 QY 2 QEFMALELNLPGEFNSGNQPSNREKRRRDILPFOH 38  
 DB 99 EEFNLSEVGPSTSEGDKNHTSKNRTNLPVNH 135  
 RESULT 9  
 PP1L\_DROME STANDARD; PRT; 1631 AA.  
 ID PP1L\_DROME  
 AC P35992;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase 10D precursor (EC 3.1.3.48) (Receptor-linked protein-tyrosine phosphatase 10D).  
 GN PTP10D.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RX TISSUE-Embryo;  
 RC MEDLINE-92034989; PubMed-1657402;  
 RA Tian S.-S., Tsoulfas P., Zinn K.;  
 RT Three receptor-linked protein-tyrosine phosphatases are selectively expressed on central nervous system axons in the Drosophila embryo. \*;  
 RT Cell 67:675-685(1991).  
 RL [2]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE-Embryo;  
 RX MEDLINE-92034988; PubMed-1657401;  
 RA Yang X., Seow K.T., Bahrt S.M., Oon S.H., Chia W.;  
 RT Two Drosophila receptor-like tyrosine phosphatase genes are expressed in a subset of developing axons and pioneer neurons in the embryonic CNS. \*;  
 RT Cell 67:661-673(1991).  
 RL -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a short form; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND PIONEER NEURONS IN THE EMBRYO.  
 CC -1- SIMILARITY: CONTAINS 12 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

DR EMBL; M80538; AAA28952.1; -  
EMBL; M80455; -  
DR EMBL; M80538; AAA28952.1; -  
EMBL; M80455; -

CARBOHYD	1195	N-LINKED (GLCNAC. . .) (POTENTIAL).
	1195	N-LINKED (GLCNAC. . .) (POTENTIAL).

- FUNCTION: SPECIFICALLY BINDS POLY(G) RNA HOMOPOLYMERS IN VITRO
- SUBCELLULAR LOCATION: CYTOSOL

- CELLULAR LOCATION: Nuclear (Probable).  
- TISSUE SPECIFICITY: UBQUITOUS IN ADULTS.

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -1- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 CC EMBL: AF042857; AAC05826.1; -  
 CC EMBL: AF069517; AAC21578.1; -  
 CC EMBL: AF091264; AAC04160.1; -  
 CC EMBL: U50839; AAC35207.1; -  
 CC Genbank: HGNC:9903; RBM6.  
 CC MIM: 606886; -  
 CC InterPro: IPR000467; G\_patch.  
 CC InterPro: IPR000504; RNA\_rec\_mot.  
 CC InterPro: IPR003955; RRM\_2.  
 CC InterPro: IPR000822; Znf\_C2H2.  
 CC Pfam: PF01585; G\_patch; 1.  
 CC SMART: SM00443; G\_patch; 1.  
 CC SMART: SM00360; RRM; 1.  
 CC SMART: SM00362; RRM\_2; 1.  
 CC SMART: SM00355; Znf\_C2H2; 1.  
 CC PROSITE: PS50174; G\_PATCH; 1.  
 CC PROSITE: PS50102; RRM; 1.  
 CC PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 CC RMA-binding; Nuclear protein.  
 CC RMA-binding; Nuclear protein.  
 CC DOMAIN 456 536  
 CC FT 1051 1097 G-PATCH.  
 CC FT 826 829 POLY-GLU.  
 CC FT 892 895 POLY-PRO.  
 CC FT 915 921 POLY-GLU.  
 CC FT 642 642 K -> R (IN REF. 1 AND 4).  
 CC FT 710 710 A -> V (IN REF. 1).  
 CC FT 796 796 T -> S (IN REF. 1).  
 CC FT CONFLICT  
 CC SEQUENCE 1123 AA; 128615 MW; 450F9D7AF9C3FCB7 CRC64;  
 Query Match 19.9%; Score 73.5; DB 1; Length 1123;  
 Best Local Similarity 37.7%; Pred. No. 0.41; Indels 11; Gaps 2;  
 Matches 20; Conservative 7; Mismatches 15; Gaps 2;  
 Db 9 LKALPGEFSGNPNRKKRYRDLFPQHGHSQPN-----ERTTWHGSGNEG 57  
 35 LKSHQERHSGNPG-----RDSLPRFDGSHSGPPFANVEHSHFSYGARDG 80  
 RESULT 11  
 CSM\_DROME STANDARD; PRT: 845 AA.  
 ID CSM\_DROME STANDARD; PRT: 845 AA.  
 AC P29349; Q24032; Q9W524; Q24033; Q9V3H1;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase corkscrew (EC 3.1.3.48).  
 GN CSM OR EG:BACN25G24.2 OR CG3954.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM Y1229), FUNCTION, TISSUE SPECIFICITY, AND  
 RP DEVELOPMENTAL STAGE.  
 RC TISSUE-Embryo;  
 RX MEDLINE-92346711; PubMed-1638629;  
 RA Perkins L.A., Larsen I., Perrimon N.;  
 RT "Corkscrew encodes a putative protein tyrosine phosphatase that  
 RT functions to transduce the terminal signal from the receptor tyrosine  
 RT kinase torso."  
 RL Cell 70:225-236(1992).

RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS Y1229 AND 4A).  
 RC STRAIN-DP CN BW.  
 RA Melnick M.B., Melnick C.B., Larsen I., Perrimon N., Perkins L.A.;  
 RT "The role of the Drosophila corkscrew protein as a transducer  
 RT downstream of receptor tyrosine kinases is functionally conserved."  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS Y1229; 2 AND 4A).  
 RC STRAIN-Berkley;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Gelinko S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer J.R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Abmayy A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Broxton P., Brotter P.,  
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dudlin K.J., Evangelista C.C., Ferraz C., Gerhart W.M., Glaeser K.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostali M., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrel A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye Y., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORMS 4A AND Y1229).  
 RC STRAIN-Oregon-R;  
 RX MEDLINE-20196011; PubMed-10731137;  
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,  
 RA Bartell B.G., Ferraz C., Vidal S., Brun C., Demillies J., Cadieu E.,  
 RA Dreano S., Gloux S., Lelaure V., Mottler S., Gaibler F., Borokova D.,  
 RA Minana B., Karafos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
 RA Papagiannakis G., Spanos L., Cox S., Madano E., de Pablos B.,  
 RA Modelle J., Peter A., Schoettler P., Werner M., Mourikoti F.,  
 RA Belhert N., Dowe G., Scheffer U., Jaecle H., Bucheton A.,  
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,  
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
 RA Glover D.M.;  
 RT "From sequence to chromosome: the tip of the X chromosome of D.  
 RT melanogaster."  
 RL Science 287:2220-2222(2000).  
 RN [5]  
 RP FUNCTION.  
 RX MEDLINE-97105827; PubMed-8948575;  
 RA Perkins L.A., Johnson M.R., Melnick M.B., Perrimon N.;  
 RT "The nonreceptor protein tyrosine phosphatase corkscrew functions in  
 RT multiple receptor tyrosine kinase pathways in Drosophila."  
 RL Dev. Biol. 180:63-81(1996).

CC -1- FUNCTION: REQUIRED IN ALL RECEPTOR TYROSINE KINASE SIGNALING  
 CC PATHWAYS. FUNCTIONS DOWNSTREAM OF THE RECEPTOR TYROSINE KINASE  
 CC TOROS, ACTING IN CONCERT WITH D-RAF VIA TAILLESS. ALSO FUNCTIONS  
 CC DOMINANT OF EGFR (EPIDERMAL GROWTH FACTOR RECEPTOR) AND ETL  
 CC (FIBROBLAST GROWTH FACTOR RECEPTOR). THE SH2 DOMAIN SUGGESTS THAT  
 CC INTERACTIONS. MATERIALLY REQUIRED FOR NORMAL DETERMINATION OF CELL  
 CC FATES AT THE TERMINI OF THE EMBRYO. REQUIRED FOR CELL FATE  
 CC SPECIFICATION OF THE VENTRAL ECTODERM. IN THE DEVELOPING EMBRYONIC  
 CC CNS AND FOR EMBRYONIC TRACHEAL CELL MIGRATION. FUNCTIONS DURING  
 CC AS EYES, ARISTAE, L5 WING VEIN AND THE TARSAL CLAW.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O -> protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 2 (SHOWN HERE), 4A AND Y1229;  
 CC MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED UNIFORMLY THROUGHOUT ALL TISSUES  
 CC DURING EMBRYOGENESIS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT.  
 CC WHICH SHARES NO HOMOLOGIES WITH OTHER PTASE PROTEINS. THIS PTASE  
 CC INSERT IS REMINISCENT OF THE KINASE INSERT WITHIN THE KINASE  
 CC CATALYTIC DOMAINS OF SEVERAL RECEPTOR TYROSINE KINASES.  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M94730; AA28453.1; -  
 CC EMBL: U19909; AA80254.1; -  
 CC EMBL: U19909; AA80254.1; -  
 CC EMBL: AE003423; AA45724.1; -  
 CC EMBL: AE003423; AA45724.1; -  
 CC EMBL: AE003423; AA45724.1; -  
 CC EMBL: AL32797; CAB65870.1; -  
 CC EMBL: AL32797; CAB65870.1; -  
 CC EMBL: AL32797; CAB65870.1; -  
 CC PIR: A43254; A43254.  
 CC HSP: 06124; 258P.  
 CC FLYBASE: FBgn0000382; csw.  
 CC InterPro: IPR003595; PTPC\_motif.  
 CC InterPro: IPR000980; SH2.  
 CC InterPro: IPR000387; TYR\_phosphatase.  
 CC InterPro: IPR000242; Tyr\_PP.  
 CC Pfam: PF00102; SH2; 2.  
 CC Pfam: PF00102; SH2; 2.  
 CC PRINTS: PR00700; PRTYPHPTASE.  
 CC PRODOM: PD000093; SH2; 2.  
 CC SMART: SM00194; PTPC; 1.  
 CC SMART: SM00252; PTPC\_motif; 1.  
 CC PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
 CC PROSITE: PS50055; TYR\_PHOSPHATASE\_2; 1.  
 CC PROSITE: PS50055; TYR\_PHOSPHATASE\_2; 1.  
 CC PROSITE: PS50001; SH2; 2.  
 CC Hydrolase: SH2 domain; Repeat; Developmental protein;  
 CC Alternative splicing.  
 CC DOMAIN 6 101  
 CC FT 111 205 SH2 1.  
 CC FT 227 645 SH2 2.  
 CC FT 289 444 PROTEIN-TYROSINE PHOSPHATASE.  
 CC FT 583 583 PTASE INSERT (CYS/SER-RICH).  
 CC FT 583 583 BY SIMILARITY.  
 CC FT 1 110 MSSRMFPTISGIEAEKLEOEGFDGSLARLSSNGAF  
 CC VARSPLIC 1 110 TISVRGNEVTHIKIKNGGDFDLGGEKFTLDELVGYM  
 CC ENGLKKNQAIHLKPLICAEPTTER -> MLENKLEK

CC FT 111 205 SH2 1.  
 CC FT 227 645 SH2 2.  
 CC FT 289 444 PROTEIN-TYROSINE PHOSPHATASE.  
 CC FT 583 583 PTASE INSERT (CYS/SER-RICH).  
 CC FT 583 583 BY SIMILARITY.  
 CC FT 1 110 MSSRMFPTISGIEAEKLEOEGFDGSLARLSSNGAF  
 CC VARSPLIC 1 110 TISVRGNEVTHIKIKNGGDFDLGGEKFTLDELVGYM  
 CC ENGLKKNQAIHLKPLICAEPTTER -> MLENKLEK

Query Match 19.88; Score 73; DB 1; Length 845;  
 Best Local Similarity 59.18; Pred. No. 0.34;  
 Matches 13; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 17 NSGNPSNRKKNRYDLPPOH 38  
 245 NEGYKQENRKNRYNRLPYDH 266

RESULT 12  
 ID PTNR\_MOUSE STANDARD: PRT; 585 AA.  
 AC P35235.  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)  
 DE (Protein-tyrosine phosphatase SYP).  
 GN PRPN11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_Taxid:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:93206094; PubMed:8096088;  
 RA Feng G.-S., Hui C.-C., Pawson T.;  
 RT SH2-containing phosphotyrosine phosphatase as a target of protein-  
 RT tyrosine kinases".  
 RL Science 259:1607-1611(1993).  
 RP [2]  
 RX X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 4-103.  
 RA Lee C.-H., Komatsu D., Jacques S., Margolis B., Schlessinger J.,  
 RA Shoelson S.E., Kurlyan J.;  
 RT "Crystal structures of peptide complexes of the amino-terminal SH2  
 RT domain of the SYP tyrosine phosphatase".  
 RL Structure 2:423-438(1994).  
 CC -1- FUNCTION: THIS PTASE ACTIVITY MAY DIRECTLY LINK GROWTH FACTOR  
 CC RECEPTORS AND OTHER SIGNALING PROTEINS THROUGH PROTEIN-TYROSINE  
 CC PHOSPHORYLATION. THE SH2 REGIONS MAY INTERACT WITH OTHER CELLULAR  
 CC COMPONENTS TO MODULATE ITS OWN PHOSPHATASE ACTIVITY AGAINST  
 CC INTERACTING SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O -> protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- PTM: PHOSPHORYLATED BY TYROSINE-PROTEIN KINASES.  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L08663; -; NOT\_ANNOTATED\_CDS.

DR PIR; A46209; A46209.  
 DR PDB; 1A5A; 31-AUG-94.  
 DR PDB; 1AYB; 31-AUG-94.  
 DR PDB; 1AYC; 31-AUG-94.  
 DR PDB; 1AYD; 31-AUG-94.  
 DR MGD; MGI:99511; Pcpn11.  
 DR InterPro; IPR000980; SH2.  
 DR InterPro; IPR000387; TYR\_phosphatase.  
 DR InterPro; IPR000242; Tyr\_PP.  
 DR Pfam; PF00017; SH2\_2.  
 DR Pfam; PF00102; X-phosphatase; 1.  
 DR PRINTS; PR00700; PRTPHPPTASE.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR PRODOM; PD000093; SH2\_2.  
 DR SMART; SM00194; PTPc; 1.  
 DR SMART; SM00252; SH2\_2.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR PROSITE; PS50001; SH2\_2.  
 DR Hydrolase; SH2 domain; Repeat; Phosphorylation; 3D-structure.  
 FT DOMAIN 6 102 SH2 1.  
 FT DOMAIN 112 216 SH2 2.  
 FT DOMAIN 276 521 PROTEIN-TYROSINE PHOSPHATASE.  
 FT ACT\_SITE 463 463 BY SIMILARITY.  
 FT SEQUENCE 585 AA; 66816 MW; 6CE554F923B8F72A CRC64;

Query Match 19.5% Score 72; DB 1; Length 585;  
 Best Local Similarity 38.1%; Pred. No. 0.3;  
 Matches 16; Conservative 7; Mismatches 13; Indels 6; Gaps 2;

Gy 19 GN0PSNREKNRYDILPFQH-----HGSGPNERFTTFHGSN 55  
 Db 268 G0R0ENKNKNRYKNILPFQDTRVRLHD-GDPNEPVSDIYAN 308

RESULT 13  
 PTNB\_HUMAN STANDARD; PRT; 593 AA.  
 AC 006124;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)  
 DE (Protein-tyrosine phosphatase 2C) (PTP-2C) (PTP-1D) (SH-PTP3) (SH-PTP2) (SH-2).  
 GN PTPN11 OR PTP2C OR SHPTP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Umbilical cord;  
 RX MEDLINE-93211929; PubMed-7681589;  
 RA Ahmad S., Banville D.L., Zhao Z., Fischer E.H., Shan S.H.;  
 RT "A widely expressed human protein-tyrosine phosphatase containing src  
 RT homology 2 domains";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2197-2201(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-93206095; PubMed-7681217;  
 RA Vogel W., Lammers R., Huang J., Ullrich A.;  
 RT "Activation of a phosphotyrosine phosphatase by tyrosine  
 RT phosphorylation";  
 RL Science 259:1611-1614(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-T-cell;  
 RX MEDLINE-93106179; PubMed-1281790;  
 RA Adachi M., Sekiya M., Miyachi T., Matsuno K., Hinoda Y., Imai K.,  
 RA Yachi A.;  
 RT "Molecular cloning of a novel protein-tyrosine phosphatase SH-PTP3

RT with sequence similarity to the src-homology region 2.";  
 RL FEBS Lett. 314:335-339(1992).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94029983; PubMed-8216283;  
 RA Bastien L., Ramchandran C., Liu S., Adam M.;  
 RT "Cloning, expression and mutational analysis of SH-PTP2, human  
 RT protein-tyrosine phosphatase";  
 RL Biochem. Biophys. Res. Commun. 196:124-133(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-93087502; PubMed-1280823;  
 RA Freeman R.M., Jr., Plutsky J., Neel B.G.;  
 RT "Identification of a human src homology 2-containing protein-tyrosine-  
 RT phosphatase: a putative homolog of Drosophila corkscrew";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11239-11243(1992).  
 RN [6]  
 RP PHOSPHORYLATION BY PDGFR.  
 RX MEDLINE-94316690; PubMed-8041791;  
 RA Bennett A.M., Tang T.L., Sugimoto S., Walsh C.T., Neel B.G.;  
 RT "Protein-tyrosine phosphatase SHPTP2 couples platelet-derived growth  
 RT factor receptor beta to Ras";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7335-7339(1994).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 1-526.  
 RX MEDLINE-98150850; PubMed-9491886;  
 RA Hof P., Plutsky S., Dhe-Paganon S., Eck M.J., Shoelson S.E.;  
 RT "Crystal structure of the tyrosine phosphatase SHP-2";  
 RL Cell 92:441-450(1998).  
 CC -1- FUNCTION: THIS PTPASE ACTIVITY MAY DIRECTLY LINK GROWTH FACTOR  
 CC RECEPTORS AND OTHER SIGNALING PROTEINS THROUGH PROTEIN-TYROSINE  
 CC PHOSPHORYLATION. THE SH2 REGIONS MAY INTERACT WITH OTHER CELLULAR  
 CC COMPONENTS TO MODULATE ITS OWN PHOSPHATASE ACTIVITY AGAINST  
 CC INTERACTING SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN HUMAN TISSUES.  
 CC -1- PARTICULARLY ABUNDANT IN HEART, BRAIN, AND SKELETAL MUSCLE.  
 CC -1- PTM: PHOSPHORYLATION OF TYR RESIDUES AT THE C TERMINUS BY  
 CC PLATELET-DERIVED GROWTH FACTOR CREATES A BINDING SITE FOR THE SH2  
 CC DOMAIN OF GRB2.  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; L08807; -; NOT ANNOTATED\_CDS.  
 CC EMBL; X70766; CA50045.1; -;  
 CC EMBL; D13540; BAA02740.2; -;  
 CC EMBL; L07527; AAA17022.1; -;  
 CC EMBL; L03535; AAA36611.1; -;  
 CC PIR; J08053; J08053.  
 CC PDB; 2SHP; 16-FEB-99.  
 CC Genew; HGNC:9644; PTPN11.  
 CC MIM; 176876; -;  
 CC InterPro; IPR000980; SH2.  
 CC InterPro; IPR000387; TYR\_phosphatase.  
 CC InterPro; IPR000242; Tyr\_PP.  
 CC Pfam; PF00017; SH2\_2.  
 CC Pfam; PF00102; Y-phosphatase; 1.  
 CC PRINTS; PR00700; PRTPHPPTASE.  
 CC PRODOM; PD000093; SH2\_2.  
 CC SMART; SM00194; PTPc; 1.  
 CC SMART; SM00252; SH2\_2.

DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR PROSITE; PS50001; SH2; 2.  
 KW Hydrolyase; SH2 domain; Repeat; Phosphorylation; 3D-structure.  
 FT DOMAIN 6 102 SH2 1.  
 FT DOMAIN 112 216 SH2 2.  
 FT ACT\_SITE 276 517 PROTEIN-TYROSINE PHOSPHATASE.  
 FT MOD\_RES 542 542 PHOSPHORYLATION (BY PDGFR).  
 FT MOD\_RES 580 580 PHOSPHORYLATION (BY PDGFR).  
 FT CONFLICT 535 535 S -> R (IN REF. 3).  
 FT CONFLICT 548 548 S -> P (IN REF. 3).  
 SQ SEQUENCE 593 AA; 68011 MW; 9CDBEFFA56CCB45 CRC64;  
 Query Match 19.5%; Score 72; DB 1; Length 593;  
 Best Local Similarity 38.1%; Pred. No. 0.3;  
 Matches 16; Conservative 7; Mismatches 13; Indels 6; Gaps 2;  
 QY 19 GNQPSNREKNRYRDLIPQH-----HGYSGPNERTEFWHGSN 55  
 DB 268 GQROENKKNRYKNIILPFDHTRVYLHD-GDPNEPVSIDYINAN 308  
 ID PTNB\_RAT STANDARD; PRT; 593 AA.  
 AC P41499; Q62626;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)  
 DE (Protein-tyrosine phosphatase SYP).  
 GN PTPN11.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Sprague-Dawley;  
 RX MEDLINE=94324984; PubMed=8048963;  
 RA Ding W., Zhang W.R., Sullivan K., Hashimoto N., Goldstein B.J.;  
 RT "Identification of protein-tyrosine phosphatases prevalent in  
 RT adipocytes by molecular cloning.";  
 RL Biochem. Biophys. Res. Commun. 202:902-907(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Sprague-Dawley;  
 RX MEDLINE=94216346; PubMed=7512964;  
 RA Mel L., Dornier C.A., Huganir R.L.;  
 RT "RNA splicing regulates the activity of a SH2 domain-containing  
 RT protein tyrosine phosphatase.";  
 RL J. Biol. Chem. 269:12254-12262(1994).  
 RN [3]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=93011127; PubMed=1382983;  
 RA Hirata A., Munakata H., Hata K., Suzuki Y., Tsuiki S.;  
 RT "Purification and characterization of a rat liver protein-tyrosine  
 RT phosphatase with sequence similarity to src-homology region 2.";  
 RL Eur. J. Biochem. 209:195-206(1992).  
 RN [4]  
 RP FUNCTION: THIS PTPASE ACTIVITY MAY DIRECTLY LINK GROWTH FACTOR  
 RP RECEPTORS AND OTHER SIGNALING PROTEINS THROUGH PROTEIN-TYROSINE  
 RP PHOSPHORYLATION. THE SH2 REGIONS MAY INTERACT WITH OTHER CELLULAR  
 RP COMPONENTS TO MODULATE ITS OWN PHOSPHATASE ACTIVITY AGAINST  
 RP INTERACTING SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- PTM: PHOSPHORYLATED BY TYROSINE-PROTEIN KINASES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASSES OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
 CC -----  
 DR EMBL; U09307; AAA20543.1; -  
 DR EMBL; U05963; AAA19133.1; -  
 DR PIR; S29281; S29281.  
 DR HSSP; P35235; IATXA.  
 DR InterPro; IPR000980; SH2.  
 DR InterPro; IPR000387; TYR\_phosphatase.  
 DR InterPro; IPR000242; Tyr\_Pp.  
 DR Pfam; PF00017; SH2; 2.  
 DR Pfam; PF00102; Y\_phosphatase; 1.  
 DR PRINTS; PR00700; PTPPHRTASE.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR PRODOM; PD000093; SH2; 2.  
 DR SMART; SM00194; PTPC; 1.  
 DR SMART; SM00252; SH2; 2.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR PROSITE; PS50001; SH2; 2.  
 KW Hydrolyase; SH2 domain; Repeat; Phosphorylation.  
 FT DOMAIN 6 102 SH2 1.  
 FT DOMAIN 112 216 SH2 2.  
 FT ACT\_SITE 276 517 PROTEIN-TYROSINE PHOSPHATASE.  
 FT CONFLICT 459 459 BT SIMILARITY.  
 FT CONFLICT 75 75 A -> P (IN REF. 1).  
 FT CONFLICT 407 407 G -> GOALL (IN REF. 2).  
 FT CONFLICT 547 547 Y -> S (IN REF. 2).  
 SQ SEQUENCE 593 AA; 68033 MW; 3329F10F0F60AF48 CRC64;  
 Query Match 19.5%; Score 72; DB 1; Length 593;  
 Best Local Similarity 38.1%; Pred. No. 0.3;  
 Matches 16; Conservative 7; Mismatches 13; Indels 6; Gaps 2;  
 QY 19 GNQPSNREKNRYRDLIPQH-----HGYSGPNERTEFWHGSN 55  
 DB 268 GQROENKKNRYKNIILPFDHTRVYLHD-GDPNEPVSIDYINAN 308  
 ID PTNC\_MOUSE STANDARD; PRT; 775 AA.  
 AC P35831;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 12 (EC 3.1.3.48)  
 DE (Protein-tyrosine phosphatase P19) (P19-PTP) (MPTP-PTST).  
 GN PTPN12.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92272714; PubMed=1590786;  
 RA den Hertog J., Pals G.E., Jonk L.J., Kruller M.;  
 RT "Differential expression of a novel murine non-receptor protein  
 RT tyrosine phosphatase during differentiation of p19 embryonal  
 RT carcinoma cells.";  
 RL Biochem. Biophys. Res. Commun. 184:1241-1249(1992).  
 RN [2]  
 RP REVISIONS TO 297-416.  
 RX MEDLINE=93112015; PubMed=1472029;  
 RA Takekawa M., Itoh F., Hlnoda Y., Arimura Y., Toyota M., Sekiya M.,  
 RA Adachi M., Imai K., Yachi A.;

```

RT      *Cloning and characterization of a human cDNA encoding a novel
RT      putative cytoplasmic protein-tyrosine-phosphatase.",
RL      Biochem. Biophys. Res. Commun. 189:123-1230(1992).
RM      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BALB/c;
RX      MEDLINE-95289971; PubMed-7772023;
RA      Charrest A., Wagner J., Shen S.H., Tremblay M.L.;
RT      "Murine protein tyrosine phosphatase-PEST, a stable cytosolic protein
RL      tyrosine phosphatase.";
CC      Biochem. J. 308:425-432(1995).
CC      -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC      tyrosine + phosphate.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic
CC      -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC      TYROSINE PHOSPHATASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X63440; CAA45037.1; ALT_SEQ.
DR      EMBL; X86781; CAA60477.1; -.
DR      HSSP; 006124; 2SHP.
DR      MGD; MGI:104673; Pcpn12.
DR      InterPro; IPR000387; TYR_phosphatase.
DR      InterPro; IPR000242; TYR_PP.
DR      Pfam; PF00102; Y_phosphatase; 1.
DR      PRINTS; PR00700; PRTPHPTASE.
DR      SMART; SM00194; PTPc; 1.
DR      PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR      PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR      PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW      Hydrolyase.
FT      DOMAIN 27 291 PROTEIN-TYROSINE PHOSPHATASE.
FT      ACT_SITE 231 231 BY SIMILARITY.
FT      CONFLICT 296 296 K -> N (IN REF. 1).
FT      CONFLICT 328 332 KODSP -> DETS (IN REF. 1).
FT      CONFLICT 380 380 W -> V (IN REF. 1).
SQ      SEQUENCE 775 AA; 86992 MW; 7106D73F5014E411 CRC64;

Query Match 19.2%; Score 71; DB 1; Length 775;
Best Local Similarity 57.1%; Pred. No. 0.55;
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY      18 SGNQPSNREKNRYDILPFOH 38
      :|:|:|||||
DB      52 TGEKENYKKNRYDILPFDH 72

```

Search completed: June 6, 2003, 11:29:29  
 Job time : 11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 6, 2003, 11:27:36 ; Search time 29 Seconds  
(without alignments)  
476,040 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369  
Sequence: 1 MOEFMALELKNLPGEFNSGN.....TTFWHSNCGAVSLLRVCA 67

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	48.8	398	4	O9Y406 mus sapien
2	121	32.8	426	11	O55082 mus musculu
3	88.5	24.0	297	11	O83476 ratu
4	88.5	24.0	536	4	O8TE48 homo sapien
5	88.5	24.0	642	4	O96K06 homo sapien
6	88.5	24.0	642	11	O60986 mus musculu
7	88.5	24.0	659	11	O63477 ratu
8	88.5	24.0	659	11	O61042 mus musculu
9	88	23.8	536	11	O35128 mus musculu
10	88	23.8	595	11	O92163 mus musculu
11	88	23.8	597	11	O9R0V6 mus musculu
12	86	23.3	597	4	O969V8 homo sapien
13	86	23.3	613	11	P81718 ratu
14	86	23.3	624	4	O9UK67 homo sapien
15	85	23.0	827	5	O21055 caen
16	83	22.5	595	13	O92124 xenopus lae

17	79.5	21.5	793	11	O91V35 mus musculu
18	79.5	21.5	802	4	O96TD9 homo sapien
19	79	21.4	1064	5	O8SXB2 drosophila
20	79	21.4	1767	5	O9W4F5 drosophila
21	79	21.4	1767	5	O24495 drosophila
22	77.5	21.0	956	17	O8TPY5 methanosarc
23	76.5	20.7	289	5	O9N4M9 caen
24	75.5	20.5	833	13	O9DES7 drosophila
25	75	20.3	1647	5	O9YVW1 gallus gall
26	73.5	19.9	807	13	O91969 mus musculu
27	73.5	19.9	1112	11	O8R388 mus musculu
28	72.5	19.6	1117	11	O8R388 mus musculu
29	72	19.5	274	11	O9QW08 ratu
30	72	19.5	460	4	O96HD7 homo sapien
31	72	19.5	550	5	O44180 caen
32	72	19.5	593	13	O90687 gallus gall
33	72	19.5	597	11	O64509 mus musculu
34	71	19.2	382	11	O63745 ratu
35	71	19.2	433	13	O9PT91 brachydantio
36	70	19.0	353	4	O96AUS homo sapien
37	70	19.0	406	11	O922E7 mus musculu
38	69.5	18.8	1437	5	O44329 hiru
39	69	18.7	336	11	O9QWQ7 mus musculu
40	69	18.7	1114	4	O9H0Y6 homo sapien
41	69	18.7	1143	4	O16614 homo sapien
42	69	18.7	1291	11	O61812 mus musculu
43	69	18.7	1343	11	O64730 mus musculu
44	68.5	18.6	822	13	O91556 xenopus lae
45	68.5	18.6	1200	13	O91054 heterodontu

## ALIGNMENTS

## RESULT 1

ID O9Y406 PRELIMINARY; PRT; 398 AA.

AC O9Y406; 01-NOV-1999 (T-EMBLrel. 12, Created)  
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
DE Hypothetical 45.7 kDa protein (Fragment).  
GN DKFZ566K0524.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Ansoyge W., Wilkner U., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL050040; CAB43248.1; -  
DR HSSP; O06124; 2SHP.  
DR InterPro; IPR000387; TYR\_phosphatase.  
DR InterPro; IPR000242; TYR\_PP.  
DR Pfam; PF00102; T-phosphatase. 1.  
DR PRINTS; PR00700; PRTYPHPTASE.  
DR SMART; SM00194; PTPC; 1.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
KW Hydroxylase; Hypothetical protein.  
FT NON\_TER  
SQ SEQUENCE 398 AA; 45690 MW; 857AAD03747870A2 CRC64;

Query Match 48.8%; Score 180; DB 4; Length 398;

Best Local Similarity 94.4%; Pred. No. 1.6e-14;  
Matches 34; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MOEFMALELKNLPGEFNSGNPSNREKRRYDILPF 36  
DB 138 MOEFMALELKNLPGEFYSGNOPSNRKRRYDILPY 173



RESULT 2  
ID 055082 PRELIMINARY; PRT; 426 AA.  
AC 055082;  
DT 01-JUN-1998 (TREMblrel. 06, Created)  
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Protein-tyrosine-phosphatase (EC 3.1.3.48).  
GN PTPN20.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid-10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-TESTIS;  
RX MEDLINE-98070510; PubMed-9407093;  
RA Ohnogi M., Kuramochi S., Matsuda S., Yamamoto T.;  
RT "Molecular cloning and characterization of a novel cytoplasmic  
RT protein-tyrosine phosphatase that is specifically expressed in  
RT spermatocytes."  
RL J. Biol. Chem. 272:33092-33099 (1997).  
DR EMBL; D64141; BAA23761.1; -.  
DR HSSP; Q06124; 25HP.  
DR MGD; MGI:1196295; Ptpn20.  
DR InterPro: IPR000387; TYR-phosphatase.  
DR Pfam: PF00102; T-phosphatase; 1.  
DR PRINTS; PR00700; PRTYPHPTASE.  
DR SMART; SM00194; PTPC; 1.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
KW Hydrolase.  
SQ SEQUENCE 426 AA; 49118 MW; 2B35FB13379502F4 CRC64;  
Query Match 32.8%; Score 121; DB 11; Length 426;  
Best Local Similarity 61.1%; Pred. No. 5.6e-07;  
Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
OY 1 MOEFMALELKNLPGEFNSGNQPSNREKNRYRDLIPF 36  
DB 166 IREFLEQMTLPDDFNSGNTLQNRDKNRYRDLIPY 201  
RESULT 3  
ID 063476 PRELIMINARY; PRT; 297 AA.  
AC 063476;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Protein tyrosine phosphatase epsilon M precursor (Fragment).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid-10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SPLEEN;  
RX MEDLINE-96158952; PubMed-8579581;  
RA Nakamura K., Mizuno Y., Kikuchi K.;  
RT "Molecular cloning of a novel cytoplasmic protein tyrosine phosphatase  
RT PTPepsilon."  
RL Biochem. Biophys. Res. Commun. 218:726-732 (1996).  
DR EMBL; D78613; BAA11433.1; -.  
DR HSSP; P18052; TYPO.  
DR InterPro: IPR000242; TYR\_PP.  
DR Pfam; PF00102; Y-phosphatase; 1.  
DR PRINTS; PR00700; PRTYPHPTASE.  
DR SMART; SM00194; PTPC; 1.

DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
KW Signal.  
FT SIGNAL. 1 18  
FT NON\_TER 297 297 POTENTIAL.  
SQ SEQUENCE 297 AA; 34221 MW; BCFPA633EF6526DF CRC64;  
Query Match 24.0%; Score 88.5; DB 11; Length 297;  
Best Local Similarity 28.9%; Pred. No. 0.0049;  
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
OY 2 QEFMALELKNLPGEFNSGNQPSNREKNRYRDLIPFQH-----H 39  
DB 136 EEFNSLPSGHIOGTFELANKENREKNRYRDLIPNDHCRVILSQLDGIPCSDIYINASYID 195  
OY 40 GY-----SGPNERTT--FW 51  
DB 196 GYKEKNKFIAQGPKEIVNDFW 218  
RESULT 4  
ID 08TE48 PRELIMINARY; PRT; 536 AA.  
AC 08TE48;  
DT 01-JUN-2002 (TREMblrel. 21, Created)  
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE Tyrosine phosphatase epsilon PDI.  
GN PTPPE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid-9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wabakken T., Hauge H., Funderud S., Aasheim H.C.;  
RT "Characterisation, expression and functional aspects of a novel  
RT protein tyrosine phosphatase epsilon isoform."  
RL Submitted (FEBS-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ430580; CAD3182.1; -.  
SQ SEQUENCE 536 AA; 60799 MW; 06C29C8ACACCE6A CRC64;  
Query Match 24.0%; Score 88.5; DB 4; Length 536;  
Best Local Similarity 28.9%; Pred. No. 0.0098;  
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
OY 2 QEFMALELKNLPGEFNSGNQPSNREKNRYRDLIPFQH-----H 39  
DB 79 EEFNSLPSGHIOGTFELANKENREKNRYRDLIPNDHCRVILSQLDGIPCSDIYINASYID 138  
OY 40 GY-----SGPNERTT--FW 51  
DB 139 GYKEKNKFIAQGPKEIVNDFW 161  
RESULT 5  
ID 096K06 PRELIMINARY; PRT; 642 AA.  
AC 096K06;  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Tyrosine phosphatase epsilon.  
GN PTPPE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid-9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wabakken T.K., Hauge H., Finne E.F., Wiedlocha A., Aasheim H.C.;  
RT "Expression of human protein tyrosine phosphatase epsilon in  
RT leukocytes, a potential ERK-pathway regulating phosphatase."  
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AJ31569; CAC86583.1; -  
 DR InterPro: IPR000387; TYR\_phosphatase.  
 DR InterPro: IPR000242; TYR\_PP.  
 DR Pfam: PF00102; Y\_phosphatase\_2.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; UNKNOWN\_2.  
 DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 SO SEQUENCE 642 AA; 74581 MW; D32A28841866D948 CRC64;

Query Match 24.0%; Score 88.5; DB 4; Length 642;  
 Best Local Similarity 28.9%; Pred. No. 0.012;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

OY 2 QEFMALEKLNLPGEFNSGNQPSNREKRRYRDILPFQH-----H 39  
 DB 79 EEFNLSPGHIOGTFLANKREKRRYRNILPNDHCRVILSQLDGIQPCSDYINASYD 138

OY 40 GY-----SGPNERT--FW 51  
 DB 139 GYKEKNKFIAGPKQETVNDWF 161

## RESULT 6

ID 060986 PRELIMINARY; PRT; 642 AA.

AC 060986;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Epsilon tyrosine phosphatase cytoplasmic isoform.

GN PTPRE.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_Taxid=10090;

RE MEDLINE-96109240; PubMed-8618876;

RA Elson A.; Leder P.;

RT "Identification of a cytoplasmic, phorbol ester-inducible isoform of

protein tyrosine phosphatase epsilon."

Proc. Natl. Acad. Sci. U.S.A. 92:12235-12239(1995).

DR EMBL: U36758; AAC52331.1; -

DR HSSP: P18052; IYFO.

DR MGD: MGI:97813; PTPre.

DR InterPro: IPR000387; TYR\_phosphatase.

DR InterPro: IPR000242; TYR\_PP.

DR Pfam: PF00102; Y\_phosphatase\_2.

DR PRINTS: PR00700; PRTYPHPTASE.

DR SMART: SM00194; PTPC; 2.

DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 2.

DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 2.

DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 2.

GN Hydrolyase.

SO SEQUENCE 642 AA; 74674 MW; EB89E78259C60A47 CRC64;

Query Match 24.0%; Score 88.5; DB 11; Length 642;  
 Best Local Similarity 28.9%; Pred. No. 0.012;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

OY 2 QEFMALEKLNLPGEFNSGNQPSNREKRRYRDILPFQH-----H 39  
 DB 79 EEFNLSPGHIOGTFLANKREKRRYRNILPNDHCRVILSQLDGIQPCSDYINASYD 138

OY 40 GY-----SGPNERT--FW 51  
 DB 139 GYKEKNKFIAGPKQETVNDWF 161

## RESULT 7

ID 063477 PRELIMINARY; PRT; 659 AA.

AC 063477;

DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Protein tyrosine phosphatase epsilon C (fragment).  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPLEEN.  
 RX MEDLINE-96158952; PubMed-8579581;  
 RA Nakamura K., Mizuno Y., Kikuchi K.;  
 RT "Molecular cloning of a novel cytoplasmic protein tyrosine phosphatase  
 PTPepsilon."  
 RL Blochem. Biophys. Res. Commun. 218:736-732(1996).  
 DR EMBL: D78610; BAA20333.1; -  
 DR HSSP: P18052; IYFO.  
 DR InterPro: IPR000387; TYR\_phosphatase.  
 DR InterPro: IPR000242; TYR\_PP.  
 DR Pfam: PF00102; Y\_phosphatase\_2.  
 DR PRINTS: PR00700; PRTYPHPTASE.  
 DR SMART: SM00194; PTPC; 2.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 GN Hydrolyase.  
 FT NON\_TER  
 SO SEQUENCE 659 AA; 76451 MW; FCC459BB5DD11A4 CRC64;

Query Match 24.0%; Score 88.5; DB 11; Length 659;  
 Best Local Similarity 28.9%; Pred. No. 0.012;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

OY 2 QEFMALEKLNLPGEFNSGNQPSNREKRRYRDILPFQH-----H 39  
 DB 96 EEFNLSPGHIOGTFLANKREKRRYRNILPNDHCRVILSQLDGIQPCSDYINASYD 155

OY 40 GY-----SGPNERT--FW 51  
 DB 156 GYKEKNKFIAGPKQETVNDWF 178

## RESULT 8

ID 061042 PRELIMINARY; PRT; 699 AA.

AC 061042;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Protein tyrosine phosphatase.

GN PTPRE.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_Taxid=10090;

RE MEDLINE-96181534; PubMed-8610169;

RA Schmidt A., Rutledge S.J., Endo N., Ogas E., Tanaka H., Wesolowski G.,

Liu C.T., Huang Z., Ramachandaran C., Rodan S.B., Rodan G.A.;

"Protein-tyrosine phosphatase activity regulates osteoclast formation

and function: inhibition by alendronate."

Proc. Natl. Acad. Sci. U.S.A. 93:3068-3073(1996).

GN [2]

RP SEQUENCE FROM N.A.

RA Aaronson J.;

DR EMBL: U40280; AAB02190.1; -

DR HSSP: P18052; IYFO.

DR MGD: MGI:97813; PTPre.

DR InterPro: IPR000387; TYR\_phosphatase.

DR InterPro: IPR000242; TYR\_PP.

DR Pfam: PF00102; Y\_phosphatase; 2.  
DR PRINTS: PR00700; PRTPHPTASE.  
DR SMART: SM00194; PTPc; 2.  
DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 2.  
DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 2.  
DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
DR Hydrolyase.  
SQ SEQUENCE 699 AA; 80687 MW; 581F9CB881BC05B CRC64;  
Query Match 24.0%; Score 88.5; DB 11; Length 699;  
Best Local Similarity 28.9%; Pred. No. 0.013;  
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
OY 2 OEPMAL-ELKNGEFGNSGNPSNREKRRYRDILPFQH-----H 39  
DB 136 EEFNLSQKQEVKNLHORL-EGGRPENKSKNRYKNILPFQD 195  
OY 40 GY-----SGPNERT--FW 51  
DB 196 GYKRNKFTIAAGPKQETVNDPW 218  
RESULT 9  
ID 035128 PRELIMINARY; PRT; 556 AA.  
AC 035128;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE PPN6.  
GN PPN6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98112780; PubMed=9445485;  
RA Ansari-Lari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M.,  
Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W.,  
Gibbs R.A.;  
RT "Comparative sequence analysis of a gene-rich cluster at human  
chromosome 12p13 and its syntenic region in mouse chromosome 6.";  
RL Chromosome Res. 8:29-40(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Muzny D., Ansari-Lari M.A., Timms K.M., Yu W., Dugan S., Lu J.,  
Shen Y., Rowland K., Liu W., Perez L., Ding Y., Haywood M., Jain A.,  
Leal B., Logan O., Nguyen V., Savage L., Shen H., Worley K., Chen E.,  
Forcum J., Atkinson A.D., Chiu M.W., Gorrell J.H., Brundage E., Di W.,  
Chinault C., Nelson D., Gibbs R.A.;  
RT "Direct Submission."  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX Muzny D., Atkinson A.D., Adams C., Brundage E., Bunac C., Carvelli K.,  
Chacko J., Chen J., Di W., Ding Y., Dugan S., Durbin J., Forcum J.,  
Ganesh R., Garcia C., Goodman M., Gorrell J.H., Haywood M.,  
Hernandez J., Jackson L., Jin S., Kampel R., Karpaty S., Kovar C.,  
Leal B., Li Y., Licharge O., Liu W., Logan O., Lu J., Ly T.,  
Martinez C., Oswal G., Perez L., Rashid N.D., Rowland K., Savage L.,  
Scherer S.E., Shen H., Simon M., Stovall K., Timms K.M., Todd J.,  
Vo O., Williamson A., Worley K.C., Yu W., Chinault C., Nelson D.,  
Gibbs R.A.;  
RT "Direct Submission."  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP HSSP: P29350; 1GKZ.  
DR InterPro: IPR000980; SH2.  
DR InterPro: IPR000387; TYR\_phosphatase.  
DR Pfam: PF00017; SH2; 2.  
DR Pfam: PF00102; Y\_phosphatase; 1.

DR PRINTS: PR00700; PRTPHPTASE.  
DR Prodom: PD000093; SH2; 2.  
DR SMART: SM00194; PTPc; 1.  
DR SMART: SM00252; SH2; 1.  
DR PROSITE: PS50001; SH2; 1.  
DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 1.  
DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
DR Hydrolyase.  
SQ SEQUENCE 556 AA; 63179 MW; CACF025DF9D7BC2D CRC64;  
Query Match 23.8%; Score 88; DB 11; Length 556;  
Best Local Similarity 47.5%; Pred. No. 0.012;  
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;  
OY 2 OEPMAL-ELKNGEFGNSGNPSNREKRRYRDILPFQD 38  
DB 207 EEFNLSQKQEVKNLHORL-EGGRPENKSKNRYKNILPFQD 245  
RESULT 10  
ID 0921G3 PRELIMINARY; PRT; 595 AA.  
AC 0921G3;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hemopoietic cell phosphatase.  
GN HCPH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC012660; AAH12660.1;  
DR MGI: 96055; HCPH.  
DR InterPro: IPR000980; SH2.  
DR InterPro: IPR000387; TYR\_phosphatase.  
DR InterPro: IPR000242; TYR\_PP.  
DR Pfam: PF00102; Y\_phosphatase; 1.  
DR Prodom: PD000093; SH2; 2.  
DR PROSITE: PS50001; SH2; 2.  
DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; UNKNOWN\_1.  
DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 1.  
DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
SQ SEQUENCE 595 AA; 67545 MW; CFI6750D032638D2 CRC64;  
Query Match 23.8%; Score 88; DB 11; Length 595;  
Best Local Similarity 47.5%; Pred. No. 0.013;  
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;  
OY 2 OEPMAL-ELKNGEFGNSGNPSNREKRRYRDILPFQD 38  
DB 246 EEFNLSQKQEVKNLHORL-EGGRPENKSKNRYKNILPFQD 284  
RESULT 11  
ID 09ROV6 PRELIMINARY; PRT; 597 AA.  
AC 09ROV6;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE SH2 phosphatase 1.  
GN SHP-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C3H;
RA Tsui H.W., Tsui F.;
RT "Murine SHP-1";
RL Nat. Genet. 0:0-0(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C3H;
RA MEDLINE-9348302; PubMed-10419485;
RA Martin A., Tsui H.W., Shulman M.J., Iseman D., Tsui F.W.;
RT "Murine SHP-1 splice variants with altered Src homology 2 (SH2)
RT domains. Implications for the SH2-mediated intramolecular regulation
RT of SHP-1."
RL J. Biol. Chem. 274:21725-21734(1999).
DR EMBL; U65955; AAD00151.1; -.
DR EMBL; U65951; AAD00151.1; JOINED.
DR EMBL; U65952; AAD00151.1; JOINED.
DR EMBL; U65953; AAD00151.1; JOINED.
DR EMBL; U65954; AAD00151.1; JOINED.
DR HSSP; P29350; IGMZ.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_PP.
DR Pfam; PF00017; SH2; 2.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRYPPHPTASE.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 2.
DR SMART; SM00194; PRPC; 1.
DR SMART; SM00252; SH2; 2.
DR PROSITE; PS50001; SH2; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydroxylase.
SQ SEQUENCE 597 AA; 67717 MW; E8491CE7E06E989 CRC64;

Query Match 23.3%; Score 88; DB 11; Length 597;
Best Local Similarity 47.5%; Pred. No. 0.013;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 QEFMAL---ELKNLPGEFNSGNPSNEKNRYDILPFQ 38
DB 248 EEFESLQKQEVKNLHQL-EGQREPKSKNRYKILPFDH 286

RESULT 12
ID Q969V8 PRELIMINARY; PRT; 597 AA.
AC Q969V8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Protein tyrosine phosphatase, non-receptor type 6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007667; AAH07667.1; -.
DR EMBL; BC007623; AAH07623.1; -.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR000387; TYR_phosphatase.

```

```

DR InterPro: IPR000242; TYR_PP.
DR Pfam; PF00017; SH2; 2.
DR Pfam; PF00102; Y_phosphatase; 1.
DR ProDom; PD000093; SH2; 2.
DR PROSITE; PS50001; SH2; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Receptor.
SQ SEQUENCE 597 AA; 67719 MW; 6A291A2860159389 CRC64;

Query Match 23.3%; Score 86; DB 4; Length 597;
Best Local Similarity 47.5%; Pred. No. 0.023;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 QEFMAL---ELKNLPGEFNSGNPSNEKNRYDILPFQ 38
DB 248 EEFESLQKQEVKNLHQL-EGQREPKSKNRYKILPFDH 286

RESULT 13
ID P81718 PRELIMINARY; PRT; 613 AA.
AC P81718;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-tyrosine phosphatase N6 (EC 3.1.3.48).
GN SHP-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Aoki N., Yamaguchi-Aoki Y., Ullrich A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A KEY ROLE IN HEMATOPOIESIS. THIS PHOSPHATASE
CC MAY DIRECTLY LINK GROWTH FACTOR RECEPTORS AND OTHER SIGNALING
CC PROTEINS THROUGH PROTEIN-TYROSINE PHOSPHORYLATION. THE SH2 REGIONS
CC MAY INTERACT WITH OTHER CELLULAR COMPONENTS TO MODULATE ITS OWN
CC PHOSPHATASE ACTIVITY AGAINST INTERACTING SUBSTRATES (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN
CC TYROSINE + ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.
DR EMBL; U77038; AAD00262.1; -.
DR HSSP; P29350; IGMZ.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_PP.
DR Pfam; PF00017; SH2; 2.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRYPPHPTASE.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 2.
DR SMART; SM00194; PRPC; 1.
DR SMART; SM00252; SH2; 2.
DR PROSITE; PS50001; SH2; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydroxylase; SH2 domain.
FT DOMAIN 6 102 SH2.
FT DOMAIN 271 516 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 455 455 BY SIMILARITY.
SQ SEQUENCE 613 AA; 69578 MW; 29364B28BF45C87 CRC64;

Query Match 23.3%; Score 86; DB 11; Length 613;

```

Best Local Similarity 47.5%; Pred. No. 0.024;  
Matches 19; Conservative 6; Mismatches 11; Indels 4; Gaps 2;

OY 2 QEFMAL---ELKNLPGEFNSGNPSNREKRYRDILPFQH 38  
DB 246 EEFESLQKQEVKNLHQRL-EGORPENKSKRYKNILPFQH 286

## RESULT 14

ID 021055 PRELIMINARY; PRT: 624 AA.

AC 09UK67.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 70 kDa SHP-1L protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OC NCBI\_TaxID=9606;

DR [1]

DR MEDLINE-99428497; PubMed-10497187;

DR J. Biol. Chem. 274:28301-28307(1999).

DR HSSP; P29350; 1PFR.

DR InterPro: IPR000980; SH2.

DR InterPro: IPR000387; TYR\_phosphatase.

DR Pfam; PF00017; SH2; 2.

DR PRINTS; PR00700; PRYPPHPTASE.

DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.

DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.

DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.

DR Hydrolyase.

DR SEQUENCE 624 AA; 70117 MW; CB73D7EC8396DDA0 CRC64;

## RESULT 15

ID 021055 PRELIMINARY; PRT: 827 AA.

AC 021055; 019847;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE PTP-2 protein (EC 3.1.3.48)

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OC NCBI\_TaxID=6239;

DR [1]

DR SEQUENCE FROM N.A.

DR STRAIN-BRISTOL N2;

DR Latreille P.;

DR Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;

RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.

RN [3]

RP SEQUENCE OF 160-827 FROM N.A.

RA Gutch M.J., Tongs N.K., Hengartner M.O.;

RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL; U53332; AKK3131.1; -.

DR EMBL; AF015882; AAC21678.1; -.

DR HSSP; Q06124; 2SHP.

DR InterPro: IPR002908; Frataxin\_like.

DR InterPro: IPR000980; SH2.

DR InterPro: IPR000387; TYR\_phosphatase.

DR Pfam; PF01491; Frataxin\_Cyay; 1.

DR Pfam; PF00017; SH2; 2.

DR PRINTS; PR00700; PRYPPHPTASE.

DR PRODOM; PD000093; SH2; 2.

DR SMART; SM00194; PTPc; 1.

DR PROSITE; PS01344; FRATAXIN\_1; UNKNOWN\_1.

DR PROSITE; PS50810; FRATAXIN\_2; 1.

DR PROSITE; PS50001; SH2; 2.

DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.

DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.

DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.

DR Hydrolyase.

DR SEQUENCE 827 AA; 94923 MW; 1ED2DAD4C54024D6 CRC64;

## Query Match

Best Local Similarity 23.0%; Score 85; DB 5; Length 827;

Matches 18; Conservative 6; Mismatches 13; Indels 4; Gaps 1;

OY 2 QEFMALELKNLPGE---FNSGNPSNREKRYRDILPFQH 38  
DB 425 EEFDRLSQELPAPQYLSKREGRPVNAEKRYKNIVPFDH 465

Search completed: June 6, 2003, 11:30:06  
Job time: 31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 6, 2003, 11:28:11 ; Search time 15 Seconds

(without alignments)  
429,400 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369  
Sequence: 1 MOEFMALELNKLPEFGNSGN.....TFMHSNGAVSLIRYCA 67

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	48.8	398	2	T08716
2	88.5	24.0	699	2	JC6132
3	88.5	24.0	700	1	S12053
4	88	23.8	595	1	A44390
5	86	23.3	595	1	S20825
6	85	23.0	668	2	T34317
7	83	22.5	595	1	A55651
8	79.5	21.5	802	1	A36065
9	79	21.4	1615	2	B49502
10	79	21.4	1767	2	A49502
11	77.5	21.0	796	1	JC1285
12	75	20.9	521	1	A44267
13	75	20.3	1557	2	D41214
14	73	20.3	1630	2	C41214
15	73	19.8	841	2	A43254
16	72	19.5	550	2	T32543
17	72	19.5	585	2	A46209
18	72	19.5	593	1	JN0805
19	72	19.5	593	2	JC5167
20	72	19.5	597	1	A53593
21	71	19.2	382	1	S48748
22	71	19.2	773	1	JH0609
23	71	19.2	775	2	S55345
24	71	19.2	780	1	JC1368
25	71	19.2	802	1	B44390
26	70	19.0	363	1	S14294
27	70	19.0	387	1	A38191
28	70	19.0	387	1	A60345
29	70	19.0	415	1	A33899

30	69.5	18.8	1437	2	T31093	probable protein-t
31	69	18.7	1175	2	S51005	protein-tyrosine-p
32	69	18.7	1176	2	I58345	protein tyrosine p
33	69	18.7	1291	1	A28334	protein-tyrosine-p
34	69	18.7	1304	1	A46546	leukocyte common a
35	68.5	18.6	989	2	S69711	hypothetical prote
36	68.5	18.6	1200	2	T43148	probable protein-t
37	68	18.4	432	1	A44845	protein-tyrosine-p
38	68	18.4	435	1	TPH0N1	protein-tyrosine-p
39	68	18.4	1174	2	I38140	protein-tyrosine-p
40	68	18.4	1237	2	A54080	protein-tyrosine-p
41	67	18.2	356	2	UW0049	protein-tyrosine-p
42	67	18.2	829	1	A47373	protein-tyrosine-p
43	66.5	18.0	405	2	I49372	protein-tyrosine-p
44	66.5	18.0	405	2	S68250	protein-tyrosine-p
45	66.5	18.0	1187	1	A53661	protein-tyrosine-p

## ALIGNMENTS

RESULT 1  
T08716  
protein-tyrosine-phosphatase homolog DKFZp566K0524.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999  
C:Accession: T08716  
R:Ansoorge, W.; Winkler, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16472  
A:Accession: T08716  
A:Molecule type: mRNA  
A:Residues: 1-398 <ANS>  
A:Cross-references: EMBL:AL050040  
A:Experimental source: fetal kidney; clone DKFZp566K0524  
C:Genetics:  
A>Note: DKFZp566K0524.1  
C:Superfamily: protein-tyrosine-phosphatase homology  
F:161-379/Domain: protein-tyrosine-phosphatase homology <P>

Query Match 48.8%, Score 180; DB 2; Length 398;  
Best Local Similarity 94.4%; Pred. No. 3.6e-14;  
Matches 34; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MOEFMALELNKLPEFGNSGNPSNREKRYDILPF 36  
|||||  
Db 138 MOEFMALELNKLPEFGNSGNPSNREKRYDILPF 173

## RESULT 2

JC6132  
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon precursor - mouse  
N:Alternate names: phosphotyrosine phosphatase; protein-tyrosine-phosphate phosphohyd  
C:Date: 16-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 21-Jun-2002  
C:Accession: JC6132  
R:Schmidt, A.; Rutledge, S.J.; Endo, N.; Ogas, E.E.; Tanaka, H.; Wesolowski, G.; Leu,  
Proc. Natl. Acad. Sci. U.S.A. 93, 3068-3073, 1996  
A:Title: Protein-tyrosine phosphatase activity regulates osteoclast formation and fun  
A:Reference number: JC6132; MUID:96181534; PMID:8610169  
A:Contents: bone marrow cell  
A:Accession: JC6132  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-699 <SCH>  
A:Cross-references: GB:U40280; NID:91373052; PIDN:AA02190.1; PID:91373053  
C:Comment: This enzyme plays an important role in osteoclast formation and function 1  
C:Genetics:  
A:Gene: ptp  
C:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common an  
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; ty  
F:77-697/Domain: leukocyte common antigen cytosolic domain homology <LAC>

F:158-382/Domain: protein-tyrosine-phosphatase homology <PRP1>  
 F:334/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:340/Binding site: substrate phosphate (Arg) #status predicted  
 F:629/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:635/Binding site: substrate phosphate (Arg) #status predicted

Query Match 24.0% Score 88.5; DB 2; Length 699;  
 Best Local Similarity 28.9%; Pred. No. 0.011;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

QY 2 OEPMALNKLPGEFNSGNQPSNRKRRYRDLIPFQH-----H 39  
 DB 136 EEFNLSPSGHIOGTFFELANKREKRNRYPNLPNDHCRVILSQVDCIPGSDYINASYID 195  
 QY 40 GY-----SGNERTT--FW 51  
 DB 196 GYKERNKFTAAOGPKQETVNDWF 218

RESULT 3  
 S12053  
 Protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon precursor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S12053  
 R:Kruenger, N.X.; Streuli, M.; Saito, H.  
 EMBO J. 9, 3241-3252, 1990  
 A>Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatases  
 A:Reference number: S12049; MUID:91006018; PMID:2170109  
 A:Accession: S12053  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-700 <NRD>  
 A:Cross-references: GB:X54134; NID:935791; PIDN:CAA38069.1; PID:935792  
 C:Genetics:  
 A:Gene: GDB:PTPR  
 A:Cross-references: GDB:131385; OMIM:600926  
 A:Map position: 10q26-10q26  
 C:Function:  
 A:Description: catalyzes the hydrolysis of peptidyl-tyrosine to release phosphate  
 C:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antigen  
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; receptor; transmembrane prot  
 F:1-19/Domain: signal sequence #status predicted <Sig>  
 F:20-700/Product: protein-tyrosine-phosphatase, receptor type epsilon #status predicted  
 F:47-63/Domain: transmembrane #status predicted <TMN>  
 F:78-698/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
 F:159-383/Domain: protein-tyrosine-phosphatase intermediate) #status predicted  
 F:334/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:340/Binding site: substrate phosphate (Arg) #status predicted  
 F:630/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:636/Binding site: substrate phosphate (Arg) #status predicted

Query Match 24.0% Score 88.5; DB 1; Length 700;  
 Best Local Similarity 26.9%; Pred. No. 0.011;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

QY 2 OEPMALNKLPGEFNSGNQPSNRKRRYRDLIPFQH-----H 39  
 DB 137 EEFNLSPSGHIOGTFFELANKREKRNRYPNLPNDHCRVILSQVDCIPGSDYINASYID 196  
 QY 40 GY-----SGNERTT--FW 51  
 DB 197 GYKERNKFTAAOGPKQETVNDWF 219

RESULT 4  
 A44390  
 protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 6 - mouse  
 N:Alternate names: hematopoietic cell phosphatase (HCP); protein-tyrosine-phosphatase 10  
 C:Species: Mus musculus (house mouse)  
 C:Date: 03-May-1994 #sequence\_revision 19-May-1994 #text\_change 11-Jun-1999  
 C:Accession: A44390; F61180; I65741; I52816; I65740; A45143; B45143  
 R:Matthews, R.J.; Bowne, D.B.; Flores, E.; Thomas, M.L.

MOL. Cell. Biol. 12, 2396-2405, 1992  
 A>Title: Characterization of hematopoietic intracellular protein tyrosine phosphatase  
 and theonine-rich sequences.  
 A:Reference number: A44390; MUID:92236615; PMID:1373816  
 A:Accession: A44390  
 A:Molecule type: mRNA  
 A:Residues: 1-595 <MAT>  
 A:Cross-references: GB:M90389; NID:9200550; PIDN:AAA40007.1; PID:9200551  
 R:Yl, T.L.; Cleveland, J.L.; Ihle, J.N.  
 MOL. Cell. Biol. 12, 836-846, 1992  
 A>Title: Protein tyrosine phosphatase containing SH2 domains: characterization, prefe  
 A:Reference number: A42031; MUID:92123209; PMID:1732748  
 A:Accession: A42031  
 A:Molecule type: mRNA  
 A:Residues: 1-595 <YII>  
 A:Cross-references: GB:M68902; NID:9193807  
 A>Note: sequence extracted from NCBI backbone (NCBIN:76845, NCBI:76846)  
 R:Yl, T.L.; Cleveland, J.L.; Ihle, J.N.  
 Blood 78, 2222-2228, 1991  
 A>Title: Identification of novel protein tyrosine phosphatases of hematopoietic cells  
 A:Reference number: A61180; MUID:92032882; PMID:1932742  
 A:Accession: F61180  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 342-451 <YIA>  
 R:Shultz, L.D.; Schweitzer, P.A.; Rajan, T.V.; Yi, T.; Ihle, J.N.; Matthews, R.J.; Th  
 Cell 73, 1445-1454, 1993  
 A>Title: Mutations at the murine motheaten locus are within the hematopoietic cell pr  
 A:Reference number: I52816; MUID:93313972; PMID:8324828  
 A:Accession: I65741  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 334-353, 359-382 <SHU1>  
 A:Cross-references: GB:S63803; NID:9388449  
 A>Note: deletion mutation  
 A:Accession: I52816  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 350-358, 'EGSPNLTPTPSSLVLYQYTO', 359-366 <SHU2>  
 A:Cross-references: GB:S63763; NID:9388447  
 A>Note: insertion mutation  
 A:Accession: I65740  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 34-76, 'VPRPHWRAGGVTAAQGRALD' <SHU3>  
 A:Cross-references: GB:S63764; NID:9388450  
 A>Note: frameshift mutation  
 R:Yeung, Y.G.; Berg, K.L.; Pixley, F.J.; Angeletti, R.H.; Stanley, E.R.  
 J. Biol. Chem. 267, 23447-23450, 1992  
 A>Title: Protein tyrosine phosphatase-1C is rapidly phosphorylated in tyrosine in mac  
 A:Reference number: A45143; MUID:93054686; PMID:1385421  
 A:Accession: A45143  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 137-139, 'X', 141-143, 'X', 145-151 <YEU>  
 A:Experimental source: BAC1.2P5 macrophage  
 A>Note: sequence extracted from NCBI backbone (NCBIP:118519)  
 A:Accession: B45143  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 54-56, 'X', 58, 'X', 60-61, 'X', 63-68 <YEU>  
 A:Experimental source: BAC1.2P5 macrophage  
 A>Note: sequence extracted from NCBI backbone (NCBIP:118518)  
 C:Comment: This protein is found primarily in hematopoietic tissues.  
 C:Genetics:  
 A:Gene: me/HCPH; motheaten  
 C:Function:  
 A:Description: catalyzes hydrolysis of peptidyl-tyrosine to release phosphate  
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-pho  
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase  
 F:110-211/Domain: SH2 homology <SH2A>  
 F:265-521/Domain: SH2 homology <SH2B>  
 F:265-521/Domain: phosphatase catalytic domain #status predicted <PHP>

F:270-504/Domain: protein-tyrosine-phosphatase homology <PRP>  
 F:453/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:459/Binding site: substrate phosphate (Arg) #status predicted

Query Match 23.8% Score 86; DB 1; Length 595;  
 Best Local Similarity 47.5%; Pred. No. 0.011;

Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 QEFMAL---ELKNLPGEFNSGNQPSNRKRNRYDILPRQH 38  
 Db 246 EEFESLQKQEVKNLHQR-EGORPENKSKNRYKNIIPDH 284

# RESULT 5

protein-tyrosine-phosphatase (EC 3.1.3.48) PTPN6, nonreceptor type 6 [validated] - human  
 N:Alternate names: hematopoietic cell phosphatase HCP; protein-tyrosine-phosphatase 1C;  
 C:Species: Homo sapiens (man)  
 C:Date: 19-May-1994 #sequence-revision 08-Feb-1996 #text-change 21-Jun-2002  
 C:Accession: B42031; A38189; S20825; S17234; S20837  
 R:Y1, T.L.; Cleveland, J.L.; Ihle, J.N.  
 Moll. Cell. Biol. 12, 836-846, 1992  
 A:Title: Protein tyrosine phosphatase containing SH2 domains: characterization, preferen  
 A:Reference number: A42031; MUID:92123209; PMID:1732748  
 A:Accession: B42031

A:Molecule type: mRNA  
 A:Residues: 1-595 <Y11>

A:Cross-references: GB:M74093

A:Experimental source: T-lymphoid cell line

A:Note: sequence extracted from GenBank

R:Plutsky, J.; Neel, B.G.; Rosenberg, R.D.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 1123-1127, 1992

A:Title: Isolation of a src homology 2-containing tyrosine phosphatase.

A:Reference number: A38189; MUID:92141214; PMID:1736296

A:Accession: A38189

A:Molecule type: mRNA

A:Residues: 1-85, 'V', '87-595 <PLU>

A:Cross-references: GB:M77273; NID:q338079; PIDN:AAA6610.1; PID:q338080

A:Note: sequence extracted from NCBI backbone (NCBI:79619, NCBI:79620)

R:Shen, S.H.; Bastien, L.; Posner, B.I.; Chretien, P.  
 Nature 333, 868, 1991

A:Title: Corrigendum: A protein-tyrosine phosphatase with sequence similarity to the SH2

A:Reference number: S20825

A:Accession: S20825

A:Molecule type: mRNA

A:Residues: 1, 'LSRG', '4-595 <SHE>

A:Cross-references: EMBL:X62055; NID:q35781; PIDN:CAA43982.1; PID:q35782

R:Shen, S.H.; Bastien, L.; Posner, B.I.; Chretien, P.  
 Nature 352, 736-739, 1991

A:Title: A protein-tyrosine phosphatase with sequence similarity to the SH2 domain of th

A:Reference number: S17234; MUID:91343005; PMID:1652101

A:Accession: S17234

A:Molecule type: mRNA

A:Residues: 1, 'LSRG', '4-589, 'VPSRGSERCCPQVAMPQ' <SH2>

A:Experimental source: breast carcinoma cells

A:Note: sequence revised in reference S20805

C:Genetics:

A:Gene: GDB:PTPN6

A:Cross-references: GDB:131389; OMIM:176883

A:Map position: 12p13-12p13

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phosph

F:4-98/Domain: SH2 homology <SH2>

F:110-211/Domain: SH2 homology <SH2>

F:265-521/Domain: phosphatase catalytic domain #status predicted <PHP>

F:270-504/Domain: protein-tyrosine-phosphatase SH-PTP2 is required upstream

F:453/Active site: Cys (phosphocysteine intermediate) #status predicted

F:459/Binding site: substrate phosphate (Arg) #status predicted

Query Match

Best Local Similarity 23.3%; Score 86; DB 1; Length 595;

Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 QEFMAL---ELKNLPGEFNSGNQPSNRKRNRYDILPRQH 38  
 Db 246 EEFESLQKQEVKNLHQR-EGORPENKSKNRYKNIIPDH 284

# RESULT 6

protein-tyrosine-phosphatase (EC 3.1.3.48) PTP-2, nonreceptor type 6/11 [validated]  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence-revision 29-Oct-1999 #text-change 30-Jun-2002  
 C:Accession: T34317; T16193; T42211  
 R:Lattelle, P.  
 submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F59G1.  
 A:Reference number: Z21504

A:Accession: T34317

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-668 <LAT>

A:Cross-references: EMBL:U53332; PIDN:MAC71160.1; GSPDB:GN00020; CESP:F59G1.5

A:Experimental source: strain Bristol N2; clone F59G1

R:Lattelle, P.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid F28B12.

A:Reference number: Z18474

A:Accession: T16193

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-668 <LAT>

A:Cross-references: EMBL:U29537; NID:9868268; PID:9868272; PIDN:AAA68799.1; CESP:F28B

A:Experimental source: strain Bristol N2

R:Guth, M.J.; Flint, A.J.; Keller, J.; Tonks, N.K.; Hengartner, M.O.  
 Genes Dev. 12, 571-585, 1998

A:Title: The Caenorhabditis elegans SH2 domain-containing protein tyrosine phosphatase

A:Reference number: Z22076; MUID:98139525; PMID:9472025

A:Accession: T42211

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-668 <GUT>

A:Cross-references: EMBL:AF015882; NID:92323369; PIDN:MAC21678.1; PID:92323370

C:Genetics:

A:Gene: ptp-2; F59G1.5; F28B12.4

A:Map position: 2

A:Introns: 7/2; 230/3; 269/3; 586/3

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-pho

C:Keywords: phosphoric monoester hydrolase

Query Match 23.0% Score 85; DB 2; Length 668;  
 Best Local Similarity 43.9%; Pred. No. 0.029;

Matches 18; Conservative 6; Mismatches 13; Indels 4; Gaps 1;

QY 2 QEFMAL---ELKNLPGEFNSGNQPSNRKRNRYDILPRQH 38  
 Db 266 EEFRLSQEALPAEOYLTSKREGRRVNAEKRYKNIIPDH 306

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 11 - African clawed frog  
 N:Alternate names: SH-PTP2  
 C:Species: Xenopus laevis (African clawed frog)  
 C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 10-Sep-1999

C:Accession: A55651

R:Tung, T.L.; Freeman, Jr., R.M.; O'Reilly, A.M.; Neel, B.G.; Sokol, S.Y.  
 Cell 80, 473-483, 1995

A:Title: The SH2-containing protein-tyrosine phosphatase SH-PTP2 is required upstream

A:Reference number: A55651; MUID:95163101; PMID:7859288

A:Accession: A55651

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-595 <TAN>

A:Cross-references: GB:U15287; NID:9601781; PIDN:AAA65731.1; PID:9601782

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-pho



C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase  
F:6-100/Domain: SH2 homology <SH2A>  
F:112-214/Domain: SH2 homology <SH2B>  
F:273-510/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:459/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:465/Binding site: substrate phosphate (Arg) #status predicted

Query Match 22.5%; Score 83; DB 1; Length 595;  
Best Local Similarity 40.5%; Pred. No. 0.044;  
Matches 17; Conservative 8; Mismatches 11; Indels 6; Gaps 2;

QY 19 GNOPSREKKNRYDLPPOH-----HGSGPNERFTFMHGSN 55  
DB 268 GORENKNKNRYKNTLPDHTREVELHD-GDPNEQVSDYINAN 308

## RESULT 8

A36065  
Protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type alpha precursor - human

C:Species: Homo sapiens (man)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
A:Accession: A36065; S12049; S17371; S12905; S13085  
R:Kaplan, R.; Morse, B.; Huhner, K.; Croce, C.; Hawk, R.; Ravera, M.; Ricca, G.; Jaye, M.; Natl. Acad. Sci. U.S.A. 87, 7000-7004, 1990  
A:Title: Cloning of three human tyrosine phosphatases reveals a multigene family of receptors  
A:Reference number: A36065; MUID:90384936; PMID:2169617  
A:Accession: A36065

A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-802 <KAP>

A:Cross-references: GB:A3668; NID:9190738; PIDN:AAA36528.1; PID:9190739  
R:Kruener, M.X.; Streuli, M.; Salto, H.  
EMBO J. 9, 3241-3252, 1990

A:Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatases  
A:Reference number: S12049; MUID:9106018; PMID:2170109  
A:Accession: S12049

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-138,148-802 <KRU>

A:Cross-references: GB:A54150; NID:935785; PIDN:CAA38065.1; PID:935786  
R:Jirik, F.R.; Janzen, N.M.; Melhado, I.G.; Harder, K.W.; Leslie, K.B.  
submitted to the EMBL Data Library, June 1990

A:Description: Isolation of a cDNA encoding a novel protein-tyrosine phosphatase from thymocytes  
A:Reference number: S17371  
A:Accession: S17371

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-121, 'P', 123-138, 148-199, 'G', 201-203, 'C', 205-802 <JIR>

A:Cross-references: EMBL:X53364  
R:Jirik, F.R.; Janzen, N.M.; Melhado, I.G.; Harder, K.W.  
FEBS Lett. 273, 239-242, 1990

A:Title: Cloning and chromosomal assignment of a widely expressed human receptor-like protein-tyrosine phosphatase  
A:Reference number: S12905; MUID:91032191; PMID:2172030  
A:Accession: S12905

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-61, 'V', 62-82, 84-121, 'P', 123-138, 148-199, 'G', 201-203, 'C', 205-802 <JIR>  
R:Obach, S.; Nishii, M.; Steiner, D.F.  
Nucleic Acids Res. 18, 7159, 1990

A:Title: Sequence of a cDNA encoding human LRP (leukocyte common antigen-related peptide)  
A:Reference number: S13085; MUID:91088320; PMID:2175890  
A:Accession: S13085

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-113, 'N', 115-138, 148-288, 'E', 290-366, 'A', 368-492, 'S', 494-785, 'E', 787-802 <C>

A:Cross-references: EMBL:X54890; NID:932312; PIDN:CAA38662.1; PID:932313  
A:Accession: S13085

A:Gene: GDB:PTPRA; PTPA  
A:Cross-references: GDB:126732; OMIM:176884  
A:Map position: 20p13-20p13

C:Function: catalyzes the hydrolysis of peptidyl-phosphotyrosine to release phosphate  
A:Note: allows recovery from insulin stimulation by dephosphorylating insulin receptor

C:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antigen (LCA)  
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; rec  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-802/Product: protein-tyrosine-phosphatase, receptor type alpha #status predicted  
F:20-154/Domain: extracellular #status predicted <EXT>  
F:155-171/Domain: transmembrane #status predicted <TMN>  
F:172-802/Domain: intracellular #status predicted <INT>

F:184-800/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
F:285-490/Domain: protein-tyrosine-phosphatase homology <PTP1>  
F:558-780/Domain: protein-tyrosine-phosphatase homology <PTP2>  
F:442/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:448/Binding site: substrate phosphate (Arg) #status predicted  
F:732/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:738/Binding site: substrate phosphate (Arg) #status predicted

Query Match 21.5%; Score 79.5; DB 1; Length 802;  
Best Local Similarity 26.5%; Pred. No. 0.17;  
Matches 22; Conservative 9; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFMALELNLPDEFNGNPSREKKNRYDLPPOH-----H 39  
DB 243 EEFNALPACPIQATCEAASKEENKKNRYNLLPYDSRVHLTPVEGVPSDYINAFIN 302

QY 40 GY-----SGPNERFT--FW 51  
DB 303 GYQEKKNKFLAAGCPKERTVDFW 325

## RESULT 9

B49502  
Protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type 4E, splice form B precursor  
C:Species: Drosophila melanogaster  
C>Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 24-Apr-1998

A:Accession: B49502  
R:Oon, S.H.; Hong, A.; Yang, X.; Chia, W.  
J. Biol. Chem. 268, 23964-23971, 1993

A:Title: Alternative splicing in a novel tyrosine phosphatase gene (DPTP4E) of Drosophila  
A:Reference number: B49502; MUID:94043220; PMID:8226938  
A:Accession: B49502

A:Molecule type: mRNA  
A:Residues: 1-1615 <OON>  
A:Cross-references: GB:I120894

C:Genetics:  
A:Gene: FlyBase:PTP4E  
A:Cross-references: FlyBase:Fbgn0004368  
A:Introns: 1605/3

C:Superfamily: protein-tyrosine-phosphatase, receptor type 4E; fibronectin type III r  
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; rec  
F:1254-1270/Domain: transmembrane #status predicted <TMN>  
F:1271-1615/Domain: intracellular #status predicted <INT>  
F:1353-1573/Domain: protein-tyrosine-phosphatase homology <PTP1>  
F:1525/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:1531/Binding site: substrate phosphate (Arg) #status predicted

Query Match 21.4%; Score 79; DB 2; Length 1615;  
Best Local Similarity 25.3%; Pred. No. 0.44;  
Matches 25; Conservative 10; Mismatches 28; Indels 36; Gaps 3;

QY 2 QEFMALELNLPDEFNGNPSREKKNRYDLPPOH-----H 39  
DB 1331 EEFELKHHVGRQACSFALPCNRPKNRTNLLPYDSRVHLTPVEGVPSDYINAYMP 1390

QY 40 GYSGPNE-----RTFMH-----GSNEGAVSLLR 64  
DB 1391 GNSPREFIVTGGPHTSTREERWBMKMESSRAIVLTR 1429

## RESULT 10

A49502  
Protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 4E, splice form A precursor  
C:Species: Drosophila melanogaster  
C>Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 24-Apr-1998  
A:Accession: A49502





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 11:28:36 ; Search time 14 Seconds  
(Without alignments)  
140.810 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369

Sequence: 1 MOEFMALEIKNLKPEFGNSGN.....TTFWHSGNAGVSLIRYCA 67

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCFUS.COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.5	24.0	699	1	US-08-348-006B-7
2	88.5	24.0	699	2	US-08-800-825A-7
3	88.5	24.0	699	4	US-09-158-657-7
4	88	23.8	513	1	US-08-202-389-2
5	86	23.3	595	1	US-08-202-389-6
6	86	23.3	631	1	US-08-202-389-8
7	79.5	21.5	793	1	US-08-015-985-3
8	79.5	21.5	802	1	US-08-015-985-3
9	78.5	21.3	155	4	US-08-821-278A-17
10	75.5	20.5	253	2	US-08-685-992-10
11	75.5	20.5	253	2	US-09-144-925-10
12	75	20.3	263	2	US-08-685-992-13
13	75	20.3	263	2	US-09-144-925-13
14	75	20.3	273	1	US-08-201-697-17
15	74	20.1	251	2	US-08-685-992-15
16	74	20.1	251	2	US-09-144-925-15
17	73.5	19.9	125	4	US-09-152-060-103
18	73.5	19.9	471	4	US-09-152-060-102
19	72	19.5	266	2	US-08-685-992-12
20	72	19.5	266	2	US-09-144-925-12
21	72	19.5	553	1	US-08-202-389-12
22	71	19.2	253	2	US-08-685-992-20
23	71	19.2	253	2	US-09-144-925-20
24	71	19.2	272	4	US-08-821-278A-19
25	71	19.2	802	4	US-09-081-345-18
26	70	19.0	593	1	US-08-018-129-5
27	70	19.0	593	2	US-08-448-250-5

28	69.5	18.8	254	2	US-08-685-992-9	Sequence 9, Appl
29	69.5	18.8	254	4	US-09-144-925-9	Sequence 9, Appl
30	69	18.7	278	4	US-08-821-278A-18	Sequence 18, Appl
31	69	18.7	292	1	US-08-036-210-12	Sequence 12, Appl
32	69	18.7	292	2	US-08-449-609-12	Sequence 12, Appl
33	68	18.4	194	2	US-08-685-992-2	Sequence 2, Appl
34	68	18.4	194	2	US-09-144-925-2	Sequence 2, Appl
35	68	18.4	196	2	US-08-685-992-1	Sequence 1, Appl
36	68	18.4	196	2	US-09-144-925-1	Sequence 1, Appl
37	68	18.4	320	2	US-08-446-345-34	Sequence 34, Appl
38	68	18.4	320	3	US-08-964-308-2	Sequence 2, Appl
39	68	18.4	320	3	US-08-964-313-2	Sequence 2, Appl
40	68	18.4	320	4	US-09-069-138-2	Sequence 2, Appl
41	68	18.4	322	3	US-08-964-308-8	Sequence 8, Appl
42	68	18.4	322	3	US-08-964-313-8	Sequence 8, Appl
43	68	18.4	322	4	US-09-069-138-8	Sequence 8, Appl
44	68	18.4	385	1	US-08-036-210-5	Sequence 5, Appl
45	68	18.4	385	2	US-08-449-609-5	Sequence 5, Appl

#### ALIGNMENTS

RESULT 1  
US-08-348-006B-7  
Sequence 7, Application US/08348006B  
Patent No. 5658756  
GENERAL INFORMATION:  
APPLICANT: RODAN, GIDEON A.  
APPLICANT: SCHMIDT, AZRIEL  
APPLICANT: RUTLEDGE, SU JANE  
TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: J. MARK HAND  
STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,006B  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/122,032  
FILING DATE: 14-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: HAND, J., MARK  
REGISTRATION NUMBER: 36,545  
REFERENCE/DOCKET NUMBER: 189921A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-3905  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 699 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-348-006B-7  
Query Match 24.0%; Score 88.5; DB 1; Length 699;  
Best Local Similarity 28.9%; Pred. No. 0.00061;  
Matches 24; Conservative 7; Mismatches 19; Gaps 3;



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIH92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 513 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-202-389-2

Query Match 23.8%; Score 86; DB 1; Length 513;  
Best Local Similarity 47.5%; Pred. No. 0.0063;  
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 OEPMAL---ELKNLPGEFNSGNQPSNREKNRYDILPFOH 38  
:||:| |:||| |:||:||||:|||||  
Db 164 EEFESLQKQEVKNLHORL-EGQRPENKSKNRKYNILPFDH 202

RESULT 5  
US-08-202-389-6  
Sequence 6, Application US/08202389  
Patent No. 5536636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutsky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Rosenberg, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
NUMBER OF SEQUENCES: 54  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIH92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIH92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 595 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-202-389-6

Query Match 23.3%; Score 86; DB 1; Length 595;  
Best Local Similarity 47.5%; Pred. No. 0.0015;  
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 OEPMAL---ELKNLPGEFNSGNQPSNREKNRYDILPFOH 38  
:||:| |:||| |:||:||||:|||||  
Db 246 EEFESLQKQEVKNLHORL-EGQRPENKSKNRKYNILPFDH 284

RESULT 6  
US-08-202-389-8  
Sequence 8, Application US/08202389  
Patent No. 5536636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutsky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Rosenberg, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
NUMBER OF SEQUENCES: 54  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIH92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:

LENGTH: 631 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-202-389-8

Query Match  
Best Local Similarity 23.3%; Score 86; DB 1; Length 631;  
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 QEFMAL---ELKNIPEGFNSGNPSNRKRYRDLFPQH  
DB 246 EEFESLQKDEYKNIHQRL-EGQRENNKGRKRYKILPFDH 284

RESULT 7  
US-08-015-985-3

Sequence 3, Application US/08015985  
Patent No. 5538886  
GENERAL INFORMATION:

APPLICANT: Schlessinger, Joseph  
APPLICANT: Sap, Jan M.  
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/015,985  
FILING DATE: 10-FEB-1993  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7683-020  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:

LENGTH: 793 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-015-985-3

Query Match  
Best Local Similarity 21.5%; Score 79.5; DB 1; Length 793;  
Matches 22; Conservative 9; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFMALELKNIPEGFNSGNPSNRKRYRDLFPQH  
DB 234 EEFNALPACPIQATCEAASKENKRNRYNILPDIHSRVHLTPVEGVPSDYINASFIN 293  
QY 40 GY-----SGPNERTT--FW 51  
DB 294 GYQENKFTIAQGFKEETVNDWF 316

RESULT 8  
US-08-015-985-1  
Sequence 1, Application US/08015985

Patent No. 5538886  
GENERAL INFORMATION:

APPLICANT: Schlessinger, Joseph  
APPLICANT: Sap, Jan M.  
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/015,985  
FILING DATE: 10-FEB-1993  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7683-020  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 802 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-015-985-1

Query Match  
Best Local Similarity 21.5%; Score 79.5; DB 1; Length 802;  
Matches 22; Conservative 9; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFMALELKNIPEGFNSGNPSNRKRYRDLFPQH  
DB 243 EEFNALPACPIQATCEAASKENKRNRYNILPDIHSRVHLTPVEGVPSDYINASFIN 302  
QY 40 GY-----SGPNERTT--FW 51  
DB 303 GYQENKFTIAQGFKEETVNDWF 325

RESULT 9  
US-08-821-278A-17

Sequence 17, Application US/08821278A  
Patent No. 6238902  
GENERAL INFORMATION:

APPLICANT: Cheng, Jili  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: Protein Tyrosine Phosphatases

FILE REFERENCE: P1010RI  
CURRENT APPLICATION NUMBER: US/08/821,278A  
CURRENT FILING DATE: 1997-03-20  
NUMBER OF SEQ ID NOS: 23  
SEQ ID NO 17

LENGTH: 155  
TYPE: PRT  
ORGANISM: Homo Sapien

US-08-821-278A-17

Query Match  
Best Local Similarity 21.3%; Score 78.5; DB 4; Length 155;  
Matches 26; Conservative 10; Mismatches 19; Indels 51; Gaps 4;





MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,992  
FILING DATE: 25-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-685-992-13

Query Match  
Best Local Similarity 20.3%; Score 75; DB 2; Length 263;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 19 GNQPSNRKRNRYDLPFOH 38  
DB 7 GOREPNKGNKRNKILPFDH 26

RESULT 13  
US-09-144-925-13  
Sequence 13, Application US/09144925  
Patent No. 5951979  
GENERAL INFORMATION:  
APPLICANT: Tonks, Nicholas  
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
STREET: Two Milltia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02421-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/144,925  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/685,992  
FILING DATE: JULY 25, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540

TELEX:  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-144-925-13

Query Match  
Best Local Similarity 20.3%; Score 75; DB 2; Length 263;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 19 GNQPSNRKRNRYDLPFOH 38  
DB 7 GOREPNKGNKRNKILPFDH 26

## RESULT 14

US-08-201-697-17  
Sequence 17, Application US/08201697  
Patent No. 5705623  
GENERAL INFORMATION:

APPLICANT: Wiggins, Roger C.  
TITLE OF INVENTION: Mammalian Glomerular Epithelial Protein  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/201,697  
FILING DATE: 25-FEB-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-UM 9783  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 273 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-201-697-17

Query Match  
Best Local Similarity 20.3%; Score 75; DB 1; Length 273;  
Matches 26; Conservative 10; Mismatches 19; Indels 42; Gaps 5;

QY 8 ELKNI-----PGEFNSGNQPSNRKRNRYDLPFOH-----GY 41  
DB 28 ELKHVGRDPCF--ADLPCKRNKRNFTNLYDSRFKIQVDDDEGSDYINANTVFGH 85

QY 42 SGPN-----RTFWH-----GSNEGAVSLLR 64  
DB 86 NSPREFVITGGLHSTRDDFWRMCSNSRAIVMLTR 122

## RESULT 15



Thu Jun 12 08:30:32 2003

us-09-935-703-7.rapb

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 11:30:11 ; Search time 45 Seconds  
(without alignments)  
153.713 Million cell updates/sec

Title: US-09-935-703-7  
Perfect score: 1 MOEPMALELNKLPGEFNSGN.....TTFMHSGNEGAVSLIRYCA 67  
Sequence: 1

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications-AA:  
1: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB pep:\*  
2: /cgn2\_6/ptodata/2/pubppa/PCT\_NEW\_PUB pep:\*  
3: /cgn2\_6/ptodata/2/pubppa/US06\_NEW\_PUB pep:\*  
4: /cgn2\_6/ptodata/2/pubppa/US06\_PUBCOMB pep:\*  
5: /cgn2\_6/ptodata/2/pubppa/US07\_NEW\_PUB pep:\*  
6: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB pep:\*  
7: /cgn2\_6/ptodata/2/pubppa/PCTUS\_PUBCOMB pep:\*  
8: /cgn2\_6/ptodata/2/pubppa/US09\_NEW\_PUB pep:\*  
9: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB pep:\*  
10: /cgn2\_6/ptodata/2/pubppa/US10\_NEW\_PUB pep:\*  
11: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB pep:\*  
12: /cgn2\_6/ptodata/2/pubppa/US60\_NEW\_PUB pep:\*  
13: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB pep:\*  
14: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	32.8	354	9	US-09-095-478-6
2	121	32.8	379	9	US-09-095-478-8
3	121	32.8	412	9	US-09-095-478-3
4	121	32.8	426	9	US-09-095-478-1
5	121	32.8	463	9	US-09-095-478-2
6	101	27.4	30	9	US-09-095-478-22
7	86	23.3	595	10	US-09-920-021A-1
8	86	23.3	843	9	US-10-072-036-117
9	86	23.3	858	9	US-10-072-036-119
10	78.5	21.3	453	9	US-10-087-993-36
11	75.5	20.5	309	10	US-09-788-626-11
12	75	20.3	322	10	US-09-788-626-14
13	74	20.1	309	10	US-09-788-626-16
14	73.5	19.9	125	9	US-09-852-797-103
15	73.5	19.9	125	10	US-09-853-161-103
16	73.5	19.9	125	10	US-09-852-659A-103
17	73.5	19.9	471	9	US-09-852-797-102
18	73.5	19.9	471	10	US-09-853-161-102
19	73.5	19.9	471	10	US-09-852-659A-102

20	72	19.5	325	10	US-09-788-626-13	Sequence 13, Appl
21	72	19.5	593	10	US-09-920-021A-3	Sequence 3, Appl
22	72	19.5	597	9	US-10-038-010-22	Sequence 22, Appl
23	71	19.2	312	10	US-09-788-626-21	Sequence 21, Appl
24	71	19.2	802	10	US-09-822-295-18	Sequence 18, Appl
25	70	19.0	180	10	US-09-925-300-1240	Sequence 1240, Ap
26	69.5	18.8	310	10	US-09-788-626-3	Sequence 10, Appl
27	68	18.4	251	10	US-09-788-626-3	Sequence 3, Appl
28	68	18.4	254	10	US-09-788-626-2	Sequence 2, Appl
29	68	18.4	435	10	US-09-986-240-1	Sequence 32, Appl
30	67.5	18.3	448	9	US-10-087-993-32	Sequence 7, Appl
31	67.5	18.3	453	9	US-10-243-687-7	Sequence 23, Appl
32	67	18.2	341	10	US-09-788-626-23	Sequence 6, Appl
33	66	17.9	264	9	US-10-245-539-6	Sequence 26, Appl
34	64.5	17.5	281	10	US-09-788-626-28	Sequence 3627, Ap
35	64	17.3	445	9	US-09-788-626-3627	Sequence 2, Appl
36	63.5	17.2	807	10	US-09-822-295-2	Sequence 55, Appl
37	62.5	16.9	1948	9	US-09-808-602-55	Sequence 45, Appl
38	62.5	16.9	1948	9	US-09-800-198-45	Sequence 26, Appl
39	62	16.8	316	10	US-09-788-626-5	Sequence 4, Appl
40	62	16.8	316	10	US-09-788-626-26	Sequence 2, Appl
41	62	16.8	317	10	US-09-788-626-4	Sequence 28, Appl
42	62	16.8	2308	12	US-10-000-954-9	Sequence 9, Appl
43	61	16.5	246	10	US-09-848-294-9	Sequence 8, Appl
44	61	16.5	250	10	US-09-848-294-8	Sequence 22, Appl
45	61	16.5	291	10	US-09-788-626-22	

ALIGNMENTS

RESULT 1  
US-09-095-478-6  
; Sequence 6, Application US/09095478  
; Publication No. US20030095970A1  
GENERAL INFORMATION:  
APPLICANT: PLOMAN, Gregory  
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE SUPT05 AND  
TITLE OF INVENTION: RELATED PRODUCTS AND  
TITLE OF INVENTION: METHODS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,478  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 224/115  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:



STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,478  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 224/115  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 426 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
US-09-095-478-1

Query Match 32.8%; Score 121; DB 9; Length 426;  
Best Local Similarity 61.1%; Pred. No. 1.4e-07;  
Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MOEFMALEKKNLPGEFNSGNQPSNREKRYRDILPF 36  
Db 166 IREFLEQMTLPDPFNSGNTLNQRDKNRYRDILPF 201

RESULT 5  
US-09-095-478-2  
Sequence 2, Application US/09095478  
Publication No. US20030095970A1  
GENERAL INFORMATION:  
APPLICANT: Ploewman, Gregory  
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE SUPP05 AND  
TITLE OF INVENTION: RELATED PRODUCTS AND  
METHODS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,478  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 224/115  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 463 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
US-09-095-478-2

Query Match 32.8%; Score 121; DB 9; Length 463;  
Best Local Similarity 61.1%; Pred. No. 1.6e-07;  
Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MOEFMALEKKNLPGEFNSGNQPSNREKRYRDILPF 36  
Db 203 IREFLEQMTLPDPFNSGNTLNQRDKNRYRDILPF 238

RESULT 6  
US-09-095-478-22  
Sequence 22, Application US/09095478  
Publication No. US20030095970A1  
GENERAL INFORMATION:  
APPLICANT: Ploewman, Gregory  
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE SUPP05 AND  
TITLE OF INVENTION: RELATED PRODUCTS AND  
METHODS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,478  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 224/115  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid





```

; Sequence 103, Application US/09852797
; Patent No. US20020172994A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003p2
; CURRENT APPLICATION NUMBER: US/09/852,797
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 103
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-797-103

```

```

Query Match      19.9%; Score 73.5; DB 9; Length 125;
Best Local Similarity 37.7%; Pred. No. 0.052;
Matches 20; Conservative 7; Mismatches 15; Indels 11; Gaps 2;

```

```

QY 9 LKNLGGEFNSGNQPSNRKRNRYDILPFQHHGYSGRN---ERTFWHGSNEG 57
||: | :||| | | | | | | | | | | | | | | | | | | | | | | |
Db 28 LKSHAQERHSGNFPG-----RDSLPLFDQGHSGPPFANVEHRSFSYGARDG 73

```

RESULT 15  
US-09-853-161-103

```

; Sequence 103, Application US/09853161
; Patent No. US20020076756A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003p3
; CURRENT APPLICATION NUMBER: US/09/853,161
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30

```

```

; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 103
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-853-161-103

```

```

Query Match      19.9%; Score 73.5; DB 10; Length 125;
Best Local Similarity 37.7%; Pred. No. 0.052;
Matches 20; Conservative 7; Mismatches 15; Indels 11; Gaps 2;

```

```

QY 9 LKNLGGEFNSGNQPSNRKRNRYDILPFQHHGYSGRN---ERTFWHGSNEG 57
||: | :||| | | | | | | | | | | | | | | | | | | | | | | |
Db 28 LKSHAQERHSGNFPG-----RDSLPLFDQGHSGPPFANVEHRSFSYGARDG 73

```

Search completed: June 6, 2003, 11:35:49  
Job time : 46 secs



Thu Jun 12 08:30:31 2003

us-09-935-703-7.p2n.rst

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 11, 2003, 12:49:47 ; Search time 1435 Seconds

(Without alignments)  
756.166 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369

Sequence: 1 MOEFMALEIKNLPGFENSGN.....TTFMHGSEGAVALRYCA 67

Scoring table: BLOSUM62  
Xgapop 10.0, Ygapext 0.5  
Xgapop 10.0, Ygapext 0.5  
Xgapop 6.0, Ygapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 809743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip  
-Q/cgn2.1/uspro.spool/us09935703/runat\_06062003\_095211.6164/app.query.fasta.1.263  
-DB=EST -QFMT=fastap -SUFFIX=pn.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09935703.ecgn.1.1.2674.6runat.06062003.095211.6164 -NCPU=6 -ICPU=3  
-NO.MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=6 -ICPU=6  
-FGAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlun:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chain.  
score greater than or equal to the score of the result D-  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	263	71.3	571	13	BI460524
2	150	40.7	775	17	BG220714
3	145	39.3	719	17	AG752725
4	123	33.3	1842	11	AK011703
5	97.5	26.4	948	10	BB614565
6	88.5	24.0	604	10	AM321295
7	88.5	24.0	812	13	BI462537
8	88.5	24.0	1103	13	BM452119
9	88.5	24.0	2474	11	AK004602
10	88.5	24.0	2633	11	BC031078
11	88	23.8	369	13	BM429337
12	88	23.8	543	12	BF077448
13	88	23.8	727	13	BG962777
14	87.5	23.7	563	13	BE755886
15	87.5	23.7	1336	14	BM920649
16	86.5	23.4	999	13	BI768443
17	86	23.3	329	12	BE819448
18	86	23.3	349	10	AM408249
19	86	23.3	411	12	BF851252
20	86	23.3	413	10	AV696093
21	86	23.3	418	10	AM402786
22	86	23.3	436	12	BF762332
23	86	23.3	614	12	BF128847
24	86	23.3	820	12	BG398566
25	86	23.3	818	12	BG340572
26	86	23.3	849	12	BG398374
27	86	23.3	863	13	BI910569
28	86	23.3	865	14	BQ711049
29	86	23.3	919	14	BQ706936
30	84.5	22.9	347	14	D24553
31	84	22.8	339	12	BE831565
32	82	22.2	698	9	AJ395870
33	81.5	22.1	823	13	BM006699
34	81.5	22.1	943	13	BM006549
35	80.5	21.8	608	12	BG799244
36	80	21.5	514	10	AV398968
37	79.5	21.5	409	9	AA305783
38	79.5	21.5	531	9	AA572365
39	79.5	21.5	552	12	BG664544
40	79.5	21.5	563	10	BE303600
41	79.5	21.5	607	9	AU133714
42	79.5	21.5	613	10	AM412296
43	79.5	21.5	624	10	BE532150
44	79.5	21.5	639	14	BM787606
45	79.5	21.5	662	10	BE306537

#### ALIGNMENTS

RESULT 1  
LOCUS BI460524 571 bp mRNA linear EST 21-AUG-2001  
DEFINITION mRNA sequence.  
ACCESSION BI460524  
VERSION BI460524.1 GI:15251180  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCES  
1 (bases 1 to 571)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Miklos Palcovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
http://image.llnl.gov  
Plate: LLAM1673 row: n column: 10  
High quality sequence stop: 569.  
Location/Qualifiers  
1. 571  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MIM\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site: 1: BamHI, Site: 2: SalI, XhoI (gtcag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.2 kb and  
normalized to 10<sup>7</sup> 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NHGRI, National  
Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 172 a 117 c 150 g 132 t  
ORIGIN

## Alignment Scores:

Pred. No.: 3.27e-25 Length: 571  
Score: 263.00 Matches: 48  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 71.27% Indels: 0  
Gaps: 0

US-09-935-703-7 (1-67) x B1460524 (1-571)

QY 1 MetGInGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20  
|||||  
DB 427 ATGCAGAAATTTATGCTTGAACCTTAAGAACTGCTGCTGATCACTCGGAAT 486  
QY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspLleuLeuProPheGlnHisGly 40  
|||||  
DB 487 CAACCAAGACAGACAGAAACAAACAGATACCGAGATATCTTCATTTCATCATCATGA 546  
QY 41 TyrSerGlyProAsnGluArgThr 48  
|||||  
DB 547 TATAGTGGCCCAATGAGAGACAA 570

RESULT 2  
LOCUS BG220714 775 bp mRNA linear EST 21-APR-2001  
DEFINITION R5740503 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG220714  
VERSION BG220714.1 GI:13746735  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Harrington, J.J. to 775)  
Cain, S., Leventhal, C., Thurnton, M., Ramachandran, R., Whitlington, J.,  
Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith,  
J., Danzig, J., and Ducar, M.  
Creation of genome-wide protein expression libraries using random  
activation of gene expression

REFERENCE  
AUTHORS

## TITLE

JOURNAL  
MEDLINE  
COMMENT

Nat. Biotechnol. 19 (5), 440-445 (2001)  
21227151  
Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scain@atersys.com  
High quality sequence stop: 549.  
Location/Qualifiers  
1. 775  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Athersys RAGE Library"  
/cell\_line="HT1080"  
/note="See 'Creation of Genome-wide Protein Expression  
Libraries using Random Activation of Gene Expression',  
Nature Biotechnology, in press. Note that even though the  
cell type indicated is HT1080, since a random activation  
method was used, these sequence tags are not necessarily  
expressed in HT1080 under normal circumstances."

BASE COUNT 213 a 144 c 164 g 254 t  
ORIGIN

## Alignment Scores:

Pred. No.: 9.36e-10 Length: 775  
Score: 150.00 Matches: 30  
Percent Similarity: 90.91% Conservative: 0  
Best Local Similarity: 90.91% Mismatches: 3  
Query Match: 40.65% Indels: 1  
Gaps: 0

US-09-935-703-7 (1-67) x BG220714 (1-775)

QY 35 PropheGlnHisGlyTyrSerGlyProAsnGluArgThrPheTrpHisGlySer 54  
|||||  
DB 1 CCTAGTCCATCATGAGATATGAGTGGCCCAATGAGAG-ACAAAGTTCTGGCATGTCTCA 59  
55 AsnGluGlyAlaValSerLeuLeuArgTyrCysAla 67  
|||||  
DB 60 AACGAAAGACAGATATCACTTTGTTACGATATGTGCT 98

RESULT 3  
LOCUS A0752725 719 bp DNA linear GSS 19-JUL-1999  
DEFINITION HS\_5568\_B1\_G06\_RPCI-11 Human Male BAC Library Homo sapiens  
ACCESSION A0752725  
VERSION A0752725.1 GI:5539883  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Mahalas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and  
Hood, L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
Contact: Mahalas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)

TITLE  
JOURNAL  
MEDLINE  
COMMENT

**FEATURES**

**SOURCE**

Alignment Scores:	
Pred. No.:	3,36e-09
Score:	145.00
Percent Similarity:	96.67%
Best Local Similarity:	93.33%
Query Match:	39,308
DB:	17
Length:	719
Matches:	28
Conservative:	1
Mismatches:	1
Indels:	0
Gaps:	0

6 AAlaEugluLeuLeuLysAsnLeuProGlyGluIupheAsnSerGlyVAsnGluProSerAsnArg 25  
 495 GCCTTAGACTTAGAGATCTGCTGTGGTTCACACTCTGGGATCATCAACCCAGCCACAGA 55

Qy	26	GLuLysAsnArgTyrArgAspIleLeuPro	35
Db	555	GAAATAATAGATCCGAGATATTCTTCCA	584

LOCUS	AK017003	1842 bp	mRNA	linear	HTC 19-JAN-2002
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:493434J18:protein tyrosine phosphatase, non-receptor type 20, full insert sequence.				

ACCESSION	AK017003
VERSION	AK017003.1
KEYWORDS	GI:12856041
SOURCE	HTC; CAP trapper.
	Mus musculus (strain:C57BL/6J)

ORGANISM

**PARANITROPHENOL**

REFERENCE	1
AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)

## REFERENCE

TITL E

prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

## REFERENCE

**AUTHORS**  
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaaki, N., Carninci, P.,  
Komuro, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,  
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishize, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakauchi, S., Ikegami, T., Kasliwal, K.,

JOURNAL      Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE      20530913  
PUBMED      11076861

## REFERENCE AUTHORS

**AUTHORS**  
 Araiwa, J., Shitagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Kawakita, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
 Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R.,  
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
 Futschmann, W., Gaasterland, T., Gissel, C., Kling, B., Kuchwa, H.,  
 Kuehl, P., Lewis, S., Matsuno, Y., Nikaido, I., Pesole, G.,  
 Quackenbush, J., Schriml, L. M., Staudl, F., Suzuki, R., Tomita, M.,  
 Wanner, L., Mashio, T., Sakai, K., Okada, T., Fununo, M., Aono, H.,  
 Baldinetti, P., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
 Canninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
 Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
 Macchiaroni, L., Mashima, U., Mazzarelli, J., Mommaerts, P., Nordone, P.,  
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Saeki, H.,  
 Sato, K., Schobach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
 Toyooka, K., Wang, K. H., Welte, C., Whitaker, C., Wilming, L.,  
 Wraysh-Boris, A., Yoshida, K., Hasegawa, Y., Kawaj, H., Kohatsu, S.,  
 and Hayashizaki, Y.

**TITLE**  
 Functional annotation of a full-length mouse cDNA collection

**JOURNAL**  
 Nature 409 (6621), 665-660 (2001)

TITLE	JOURNAL	FUNCTIONAL ANNOTATION OF A FULL-LENGTH MOUSE CDNA COLLECTION
NATURE	409 (1998)	685-690 (2001)
MEDLINE	21085660	
PUBMED	11217851	
REFERENCE	5 (bases 1 to 1842)	
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,	

Aikawa, T., Badarelli, R., Sono, H., Brownlee, M., Bull, C.,  
Carinci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,  
Hara, A., Hayatsu, N., Hill, D., Hiroto, K., Hironaka, T., Hori, F.,  
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,  
Kuribara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,  
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,  
Sato, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
Schimi, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yamanaka, I.,  
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and  
Hayashizaki, Y.

**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Fax: 81-45-503-9216)  
**COMMENT** Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGGACAGAGAAGATCCAAAGACTCTTTTTTTTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adaptor of sequence [5' GAGGACAGAGATTCCTAGATTAAATTAATCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size selected between 0.5 and 3 kb was selected before cloning. Vector: a modified Bluescript KS(+), after bulk excision from lambda DFLC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.



OY 38 -----HSHISGLY-----TyrSer 42  
 DB 737 AGGATACATCAGCATATATAGATAGATACATGAGAGAGATATTTTATTCAGCTCA 796  
 OY 43 GLYPro-Aanglu-----ArgThrThrPheThrPhisGLYSerAsnGluGlyAlaValse 60  
 DB 797 GGGCCCTTGAGATATAGAGACTTTGGCAATGCTTCGGGNNATPAT----- 842  
 OY 60 fLeuLeuLeuArgTYrCYs 66  
 DB 843 -----TGTATCTATTGC 854  
 RESULT 6  
 AM321295 604 bp mRNA linear EST 25-JAN-2000  
 LOCUS uc029g06.y1 NCI-CGAP Mam6 Mus musculus cDNA clone IMAGE:2616346 5'  
 DEFINITION similar to gb:U35368 Mus musculus epsilon tyrosine phosphatase (MOUSE);, mRNA sequence.  
 ACCESSION AM321295 GI:6750839  
 VERSION AM321295.1  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 604)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 Tumor Gene Index  
 COMMENT  
 JOURNAL  
 AUTHORS  
 TITLE  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: Jeffrey Green M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: [www-bio.lnl.gov/bdrp/image/image.html](http://www-bio.lnl.gov/bdrp/image/image.html)  
 MGI:1024446  
 Seq primer: -40RP from GIBCO  
 High quality sequence stop: 433.  
 FEATURES  
 source  
 1. .604  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:2616346"  
 /clone\_1db="NCI-CGAP\_Mam6"  
 /sex="female, virgin"  
 /tissue\_type="infiltrating ductal carcinoma"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"  
 BASE COUNT 177 a 147 c 158 g 122 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.13 Length: 604  
 Score: 88.50 Matches: 24  
 Percent Similarity: 37.35% Conservative: 7  
 Best Local Similarity: 28.92% Mismatches: 19  
 Query Match: 23.98% Indels: 33  
 DB: 10 Gaps: 3  
 US-09-935-703-7 (1-67) x AM321295 (1-604)  
 OY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21

DB 44 GAGGAGTTCATTCATTTCGCATCTGACACACATACAGACACTTTGAACTAGCA  
 OY 22 ProSerAsnArgGluLysAsnArgTYrArgAspIleLeuProPheGlnHis-----  
 DB 104 GAAGAAAACAGAGAAAAAACAGATACCCCAACATTCGCCAATGATCATTCAGAGTG 163  
 OY 39 -----His 39  
 DB 164 ATTTGAGCCAGTGTGATGAAATCCCTGCTGACTACATTAACTCTTCAATCGAT 223  
 OY 40 GLYTYr-----SerGlyProAsnGluArgThr----- 49  
 DB 224 GGCTACAAAGAAAGAACAAATTCATACGACGCTCAAGGCCCTTACAGAGACAGTGAAT 283  
 OY 50 --PheTrp 51  
 DB 284 GACTTCTGTG 292  
 RESULT 7  
 B1462537 812 bp mRNA linear EST 21-AUG-2001  
 LOCUS 603205447F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5271347 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1462537 GI:15253193  
 VERSION B1462537.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 812)  
 NIH-MGC <http://mhc.ncbi.nlm.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL  
 TITLE  
 COMMENT  
 JOURNAL  
 AUTHORS  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shireki Toshlyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://www-bio.lnl.gov/bdrp/image/image.html>  
 Plate: L14M1684 row: m column: 12  
 High quality sequence stop: 778.  
 FEATURES  
 source  
 1. .812  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5271347"  
 /clone\_1db="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescriptPR (modified pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gcgcag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size-selected for average insert size 2.2 kb and normalized to 10^5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI/NHGR), National Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 234 a 200 c 224 g 154 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.202 Length: 812  
 Score: 88.50 Matches: 24  
 Percent Similarity: 37.35% Conservative: 7  
 Best Local Similarity: 28.92% Mismatches: 19  
 Query Match: 23.98% Indels: 33  
 DB: 13 Gaps: 3

```

QY      2  GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyLysPheAsnSerGlyAsnGln  21
Db      296  GAGCAGTTCACACTCATTCGCATCTGCATGACCATACAGAAAGAACTTTGAACTGGCAAAATAAA  355
QY      22  ProSerAsnArgGluLysAsnArgTyrAlaGlyPheAlaLeuProPheGlnHis  38
Db      356  GAAAGAAAACAGAGAAAAAAACAGATATCCCAACATCTTCCCAATGACCATCTTAGGGTG  415
QY      39  -----
Db      416  ATTCTGAGCCAACTGGATGAAATTCCTGTTCACTACATCAATGCTTCCTACATAGAT  475
QY      40  GlyTyr-----SerGlyProAsnGluArgThr  49
Db      476  GGTAAACAAAGAGAAGATAATTATATAGACAGCTCAAGGTCCCAACAGAGAAACGGTTAAC  535
QY      50  ---PheTyr  51
Db      536  GACTTCTGG  544

```

[illegible]

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
Mammalia; Eultheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 1105)  
NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov), [fn.d.](mailto:fn.d.)  
Tissue Procurement: [ATCC/DC101](mailto:ATCC/DC101)  
CDNA Library Preparation: [lifetech](mailto:lifetech)  
DNA Sequencing: [arrayed](mailto:arrayed) by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: [agencourt](mailto:agencourt) Bioscience Corporation  
Found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: [LM12200](mailto:LM12200) row: c column: 07  
High quality sequence stop: 653.  
Location: [duqlife](mailto:duqlife)

	source
1.	1103 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5526462" /clone_1lb="NH_WGC_72" /tissue_type="melanotic melanoma" /lab_host="DH10B (phage-resistant)" /note="Organ: skin; Vector: pCMV-Sport6; Site.2: Salt; Cloned unidirectionally. F Average insert size 2 kb. Library consr. Technologies."
294 a	322 c      274 g      212 t      1 others

Alignment Scores:	
Pred. No.:	0 319
Score:	88.50
Percent Similarity:	37.35%
Best Local Similarity:	26.93%
Query Match:	23.98%
DB:	13
	Gaps: 3
	Mismatches: 24
	Conservative: 7
	Indels: 19
	Length: 1103

```

QY      2  GInGIuPhemEaLalauEugluLeuLySaSnLeuProGIyLuPheAenSerGIyASnGI  21
Db      286  GAGGAGTTCACACTGTCATGTCACATCTGAGCACATCAAGAACTTTGAACTGGCAAAATPAA  345
QY      22  ProSerAsnArGIuLylSaSnArGTyTArGaPILLeuProPheGlnHis  38
Db      346  GAAGAAAAACAGAGAAAAAACAGATATCCCAACATCTTCCCAATGACCATTAGAGGTG  405
QY      39  -----
Db      406  ATTCTGAGCCAACTGAGATGAATTCCTGTCAGACTACATCAATGCTTCTCTACATAGAT  465
QY      40  GLyTYr-----SerGIyProASnGIuArGIThyr  49
Db      466  GGTTCACAAGAGAGAATTAATTCATAGACACTCAAGGTCCCAACAGGAAACGGTTAAC  525
QY      50  ---PheTyr  51
Db      526  GACTTCTGG  534

```

RESULT 9	LOCUS	DEFINITION	AK004602
			AK004602
			2474 bp
			mRNA
			linear
			HTC 19-JAN-2002
			Mus musculus adult male lung cDNA, RIKEN full-length enriched
			library, clone:1200004M14;protein tyrosine phosphatase, receptor
			type_E, full insert sequence.
			AK004602
			AK004602.1 GI:12835883
			HTC; CAP trapper.
			Mus musculus (strain:C57BL/6J) adult male

ORGANISM  
Mus musculus

REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	
1	Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
2	10349636 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., et al. The human genome: a new era of genomic medicine Nat. Rev. Genet. 1, 21-32 (2000)

TITLE	JOURNAL	PAGE	YEAR
Normalization and subtraction of capriciously-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	Genome Res.	10 (10), 1617-1630	(2000)
POBMED		20499374	
POB		11042159	
REFERENCE		3	
AUTHORS		Shibata, N., Itoh, M., Aizawa, K., Nagata, S., Goto, T.,	

Kono, H., Akiyama, J., Nishi, K., Katsunaga, S., Sasaki, N., Carninci, P., Sumi, N., Ishii, Y., Nakamura, S., Katsunagi, T., Tashiro, H., Itoh, M., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Nishigaki, T., Harada, A., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kiru, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Akakawa, T., Haru, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,  
Mizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakita, I.,  
Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadoyama, K., Matsuda, H., Asuburter, M., Batylov, S., Casavola, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, D., Kochlavi, H.,  
Kuenh, P., Lewis, S., Matsuo, Y., Nikaido, I., Pegale, E.,  
Quackenbush, J., Schriml, L. M., Staab, F., Suzuki, K., Tomita, M.,



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:5295028"
/tissue_type="Testis"
/lab_host="NIH_MGC_97"
/ab_host="DH10B"
/notes="Vector: pBluescript"

BASE COUNT      702 a      694 c      681 g      556 t
ORIGIN

Alignment Scores:
Pred. No.:      1.17
Score:          88.50
Percent Similarity: 37.35%
Best Local Similarity: 28.92%
Query Match:    23.98%
DB:             11
Gaps:           3

US-09-935-703-7 (1-67) x BC031078 (1-2633)

QY      2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21
Db      690 GAGGAGTTCACACTGCTGGCCATCTGGACACATACAGGAGACTTTTGAACCTGGCAATAA 749
QY      22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis----- 38
Db      750 GAAGAAACACAGAGAAAAAAGATATCCCAACATCTCTCCCAATGACCATCTTAGGGTG 809
QY      39 -----His 39
Db      810 ATCTGAGCACTGATGATGATCCCTGTCAGACTACATCATGCTCTCTACATAGAT 869
QY      40 GlyTyr-----SerGlyProAsnGluAlyThr--- 49
Db      870 GGTACAAAGAGAGATATATCATAGACGCTCAAGGTCCTCAACAGAAACGGTTAAC 929
QY      50 ---PheTyr 51
Db      930 GACTCTCTGG 938

RESULT 11
LOCUS      BM429937      369 bp      mRNA      linear      EST 31-JAN-2002
DEFINITION 1Duo23D2.ab1 Bos taurus Duodenum #1 library Bos taurus CDNA, mRNA
ACCESSION  BM429937
VERSION     BM429937.1 GI:18451659
KEYWORDS   EST.
SOURCE     Eukaryote
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
            1 (bases 1 to 369)
            Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Senses, C.W., Gordon
            P.M.K. and Moore, S.S.
            Gene Expression Profiling of the Bovine Gastrointestinal Tract
            Unpublished (2002)
            Contact: Dr. Stephen Moore
            Beef Genomics Laboratory
            Dept of AFNS, University of Alberta
            410 Agri/Fer, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
            Tel: 780 492 0169
            Fax: 780 492 4265
            Email: smoores@afns.ualberta.ca
            Insert Length: 369 Std Error: 0.00
            POLYA-No.

FEATURES
Source      1..369
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="Bos taurus Duodenum #1 library"
            /tissue_type="Smooth muscle"

```

```

/cell_type="Simple columnar epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMR" strain"
/notes="Organ: Intestine/Duodenum; Vector: Uni-22APXR;
Site_1: EcorI; Site_2: Xho I"

BASE COUNT      105 a      94 c      110 g      60 t
ORIGIN

Alignment Scores:
Pred. No.:      0.0726
Score:          88.00
Percent Similarity: 65.00%
Best Local Similarity: 47.50%
Query Match:    23.85%
DB:             13
Gaps:           2

US-09-935-703-7 (1-67) x BM429937 (1-369)

QY      2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db      129 GAGGAGTTCAGAGCTGTCGAGAGACAGAGAGTGAATACTTGACACGCGCTG--GAA 185
QY      19 GlyAsnGluProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db      186 GGGCAGCGCCGCAAAACAGACAGCAAGAACCGTTACAGAAACATCTCTGTTGACAC 245

RESULT 12
LOCUS      BF077448      543 bp      mRNA      linear      EST 18-OCT-2000
DEFINITION 227352 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.
ACCESSION  BF077448
VERSION     BF077448.1 GI:10871278
KEYWORDS   EST.
SOURCE     Eukaryote
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
            1 (bases 1 to 543)
            Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
            Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,
            and Keeler, J.W.
            Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
            Unpublished (2000)
            Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@maill.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross-match with the -mascote 18
            and -mismatch 12 options.
            PCR PRIMERS
            FORWARD: AGGAACAGCTATGACCAT
            BACKWARD: GTTTCACAGTCACGACG
            Plate: 51 row: F column: 7
            Seq primer: ATTAGGTGACACTATAG.

FEATURES
Source      1..543
            /organism="Sus scrofa"
            /db_xref="taxon:9623"
            /clone_lib="MARC 2P1G"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
            library made from pooled tissue from testis, ovary,
            endometrium, hypothalamus, pituitary, and placenta."

Alignment Scores:
Pred. No.:      0.129
Score:          87 t
Length:         543

```



Score: 88.00 Matches: 19  
 Percent Similarity: 65.00% Conservative: 7  
 Best Local Similarity: 47.50% Mismatches: 10  
 Query Match: 23.85% Indels: 4  
 DB: 12 Gaps: 2

US-09-935-703-7 (1-67) x BF077448 (1-543)

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

Db 137 GAGGAGTTTGAAGCTGTCGAGAAAGCAGAGAGTGAAGAACTTGACACAGCGGCTG---GAA 193

QY 19 GlysAnglnProSerAsnArgGlyLysAsnArgTyrArgAspIleLeuProPheGlnHis 38

Db 194 GGGCAGCGGCGGCAAGAAACAGACAGCAAGATCGTTACAGAAACATCTTCATTCGACAC 253

RESULT 13 BG962777 727 bp mRNA linear EST 12-JUN-2001

LOCUS 602827607F1 NC1\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4982569 5',

DEFINITION mRNA sequence.

ACCESSION BG962777.1 GI:14350414

VERSION EST.

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 727)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Unpublished (1999)

Email: cgabbs@mail.nih.gov

Tissue procurement: Jeffrey E. Green, M.D.

cDNA library preparation: Life Technologies, Inc.

DNA sequencing by: The I.M.A.G.E. Consortium (LNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://lm10985.row.m column: 02

Plate: LLM10985 row: m column: 02

High quality sequence stop: 704.

Location/Qualifiers

1..727

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone\_image="4982569"

/clone\_image="NCI\_CGAP\_Co24"

/lab\_host="DH10B (71 phage-resistant)"

/note="Organ: Colon; Vector: pCMV-SPORT6; Site:1: NotI;

Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.6 Kb. Constructed by Life

Technologies. Note: this is a NCI\_CGAP library."

BASE COUNT 202 a 182 c 208 g 135 t

ORIGIN

Alignment Scores:

Pred. No.: 0.2 Length: 727

Score: 88.00 Matches: 19

Percent Similarity: 65.00% Conservative: 7

Best Local Similarity: 47.50% Mismatches: 10

Query Match: 23.85% Indels: 4

DB: 13 Gaps: 2

US-09-935-703-7 (1-67) x BG962777 (1-727)

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

Db 222 GAGGAGTTTGAAGCTGTCGAGAAAGCAGAGAGTGAAGAACTTGACACAGCGTCTG---GAA 178

QY 19 GlysAnglnProSerAsnArgGlyLysAsnArgTyrArgAspIleLeuProPheGlnHis 38

Db 279 GGGCAGCGGCGGCAAGAAACAGACAGCAAGATCGTTACAGAAACATCTTCCTT 1

RESULT 14

LOCUS BE755886

DEFINITION 209763 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BE755886

VERSION BE755886.1 GI:10169815

KEYWORDS EST.

SOURCE COW.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 563)

Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,

Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett

,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,

Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and

Keeler,J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA

libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smithemail@marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred

v0.980904.e. Vector identified by cross-match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCATGACAT

BACKWARD: GTTTCACGACGACG

Plate: 57 row: I column: 22

Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers

1..563

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_image="MARC 2BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site:1: NotI; Site:2: SalI;

Library made from pooled tissue from testis, thymus,

semilendonsus muscle, longissimus muscle, pancreas,

adrenal, and endometrium."

BASE COUNT 172 a 135 c 153 g 103 t

ORIGIN

Alignment Scores:

Pred. No.: 0.16 Length: 563

Score: 87.50 Matches: 24

Percent Similarity: 37.35% Conservative: 7

Best Local Similarity: 28.92% Mismatches: 19

Query Match: 23.71% Indels: 33

DB: 12 Gaps: 3

US-09-935-703-7 (1-67) x BE755886 (1-563)

QY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAngln 21

Db 233 GAGGAGTTCAATTCCTGCGCTGACACAGCAAGCAAGACTTTGACATGCGCATATAA 292

QY 22 ProSerAsnArgGlyLysAsnArgTyrArgAspIleLeuProPheGlnHis 38

Db 293 GAGGAGTTCAATTCCTGCGCTGACACAGCAAGCAAGACTTTGACATGACATTCAGGGTG 352

QY 39 -----His 39

```

Db      353 ATTTAAGCCAGTGCAGGAAACCCGTCAGACTACATCAACGCTTCTACATAGAT 412
QY      40 G|T|T|Y|-----Serg|Pro|a|n|g|l|u|a|r|g|T|H|r|----- 49
         |||||
Db      413 GGTATAAGAGAAAGATTAATTTCATAGCTGCTCAAGGTCCCAAGCAGAGCATGTAAT 472
QY      50 ---Phet|p 51
         |||||
Db      473 GACTTCTGG 481

RESULT 15
LOCUS   BM920649
DEFINITION 1336 bp mRNA linear EST 12-MAR-2002
AGENCOURT_6706025 NIH_MGC_115 Homo sapiens CDNA clone IMAGE:5751852
5', mRNA sequence.
ACCESSION BM920649
VERSION   BM920649.1 GI:19371028
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1336)
AUTHORS  NIH-MGC http://mgc.ncl.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
         Email: cgaabs-remail.nih.gov
         Tissue Procurement: Life Technologies, Inc.
         CDNA Library Preparation: Life Technologies, Inc.
         CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
         DNA Sequencing by: Agencourt Bioscience Corporation
         Clone distribution: MGC clone distribution information can be
         found through the I.M.A.G.E. Consortium/LNL at:
         http://image.lnl.gov
         Plate: L1AM12785 row: b column: 13
         High quality sequence start: 15
         High quality sequence stop: 544.
         Location/Qualifiers
           1..1336
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:5751852"
             /clone_1lb="NIH_MGC_115"
             /lab_host="DH10B"
             /note="Organ: pooled brain, lung, testis; Vector:
             PCWV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA
             source anonymous pool of 6 male brains, age range 23-27; 1
             male lung, age 27; and 1 male testis, age 69. Library is
             oligo-dT primed and directionally cloned (EcoRV site is
             destroyed upon cloning). Average insert size 1.8 kb,
             insert size range 1-3 kb. Library is normalized and
             enriched for full-length clones and was constructed by C.
             Gruber (Invitrogen). Research Genetics tracking code
             021. Note: this is a NIH_MGC Library."
BASE COUNT 355 a 394 c 321 g 264 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 0.581 Length: 1336
Score: 87.50 Matches: 24
Percent Similarity: 37.35% Conservative: 7
Best Local Similarity: 28.92% Mismatches: 19
Query Match: 23.71% Indels: 33
DB: 14 Gaps: 3

US-09-935-703-7 (1-67) x BM920649 (1-1336)
QY 2 G|n|g|u|P|h|e|m|e|t|a|l|e|u|g|l|u|e|u|l|y|s|a|n|l|e|u|P|ro|g|l|y|g|l|u|P|h|e|a|n|S|e|r|g|l|y|a|n|g|l| 21
Db 247 GAGGAGTTCAACATCTGACATCTGACACATACAGAACTTTGAAGTGGCAATATAA 306
QY 22 ProSerAsnArgTulysAsnArgTylArgAspIleLeuProPheGlnHis----- 38

```

```

Db      307 GAAGAAACACAGAGAAAAACAGATATCCCAACATCCCTTCCCAATGACCATTTAGGGTG 366
QY      39 -----Serg|Pro|a|n|g|l|u|a|r|g|T|H|r|----- 39
         |||||
Db      367 ATTCGAGCCAACTGATGATTCCTGTCAGACTACATCAATGCTTCTACATAGAT 426
QY      40 G|T|T|Y|-----Serg|Pro|a|n|g|l|u|a|r|g|T|H|r|----- 48
         |||||
Db      427 GGTACAAAGAGAAAGATTAATTTCATAGCAGCTCAAGGTCCCAACAGAGAAACGGATAAC 486
QY      49 Th|P|H|e|T|p 51
         |||||
Db      487 GACTTCTGG 495

```

Search completed: June 11, 2003, 14:04:35  
 Job time: 1441 secs



XX WPI: 2001-565582/63.  
 DR N-PSDB; AAH43389.  
 XX  
 PT Novel isolated protein tyrosine phosphatase polypeptide useful for  
 PT treating and preventing Alzheimer's disease, rheumatoid arthritis,  
 PT Grave's disease, arrhythmias, neoplasms, multiple sclerosis and  
 PT diabetes mellitus  
 XX  
 PS Claim 1; Page 279-80; 286pp; English.  
 XX  
 CC The sequences given in ABA47496-500 show protein tyrosine phosphatase  
 CC (PTPase) polypeptides. The PTPase polypeptides, or the cDNA encoding  
 CC them, are useful for preventing, treating or ameliorating a medical  
 CC condition in a mammalian subject. They are useful for diagnosing,  
 CC preventing or treating neural disorders (e.g. Alzheimer's disease,  
 CC Parkinson's disease), immune system disorders (e.g. Rheumatoid  
 CC arthritis, Grave's disease), muscular disorders, reproductive  
 CC disorders, pulmonary disorders, cardiovascular disorders  
 CC (e.g. arrhythmias), infectious diseases (e.g. viral hepatitis),  
 CC hyperproliferative disorders (e.g. neoplasms), diseases at the cellular  
 CC level (e.g. multiple sclerosis), endocrine disorders (e.g. diabetes  
 CC mellitus), and wound healing and epithelial cell proliferation.  
 CC They are also useful for preventing hair loss, to prevent skin aging  
 CC due to sunburn, to maintain organs before transplantation, to modulate  
 CC mammalian characteristics, to treat weight disorders, to change a  
 CC mammal's mental or physical state, or as a food additive or  
 CC preservative.  
 CC  
 XX Sequence 67 AA:  
 SO  
 Query Match 100.0%; Score 369; DB 22; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-46;  
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MOEFMALELNLPGEFNSGNPSNRKRRYDILPFOHIGSGPNERFTFMHSGNAGVS 60  
 DB 1 MOEFMALELNLPGEFNSGNPSNRKRRYDILPFOHIGSGPNERFTFMHSGNAGVS 60  
 QY 61 LLRYCA 67  
 DB 61 LLRYCA 67  
 QY 61 LLRYCA 67  
 DB 61 LLRYCA 67  
 RESULT 2  
 ID AAEI4454 standard; Protein; 420 AA.  
 XX  
 AC AAEI4454:  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human protein phosphatase-4.  
 XX  
 KW Human: protein phosphatase; PP-4; immune system disorder; AIDS; allergy;  
 KW neurological disorder; developmental disorder; Alzheimer's disease;  
 KW cell proliferative disorder; Huntington's disease; arteriosclerosis;  
 KW renal tubular acidosis; gonadal dysgenesis; cancer; adenocarcinoma;  
 KW leukaemia; transgenic animal; gene therapy.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Region 183..411  
 FT Active-site /note="Protein-tyrosine phosphatase"  
 FT 340..388  
 FT /note="tyrosine specific protein phosphatase active  
 FT site"  
 FT 362..379  
 FT /label="Transmembrane\_domain  
 FT 351..363  
 FT /note="Tyr-phosphatase"  
 XX

PN MO200196546-A2.  
 XX  
 PD 20-DEC-2001.  
 XX  
 PF 14-JUN-2001; 2001MO-US19442.  
 XX  
 PR 16-JUN-2000; 2000US-212447P.  
 PR 22-JUN-2000; 2000US-213746P.  
 PR 29-JUN-2000; 2000US-215210P.  
 PR 06-JUL-2000; 2000US-216529P.  
 PR 12-JUL-2000; 2000US-218080P.  
 PR 21-JUL-2000; 2000US-220117P.  
 XX  
 PA (INCYTE) INCYTE GENOMICS INC.  
 XX  
 PI Au-Yang J, Baughn MR, Ding L, Elliott VS, Gandhi AR, Griffin JA;  
 PI Hafalla A, Kearney L, Lee EA, Lu Y, Nguyen DB, Patterson C;  
 PI Ramkumar J, Reddy R, Sanjanwala MS, Stewart EA, Tang YF;  
 PI Thornton M, Tribouley CM, Walla NK, Yang J, Yao MG, Yue H;  
 XX  
 DR WPI: 2002-092026/12.  
 DR N-PSDB; AAD24022.  
 XX  
 PT Novel polypeptide, useful for diagnosing, treating or preventing  
 PT disorders of growth and development, immune system, neurological and  
 PT cell proliferation diseases, comprises cancer protein phosphatase  
 PT polypeptides  
 XX  
 PS Claim 1; Page 105-106; 116pp; English.  
 XX

The present sequence is human protein phosphatase (PP)-4.  
 CC PP polynucleotide and polypeptide are useful in the diagnosis,  
 CC treatment and prevention of immune system disorders, neurological  
 CC disorders, developmental disorders and cell proliferative disorders.  
 CC Examples of immune system disorders include acquired immune deficiency  
 CC syndrome (AIDS), severe combined immunodeficiency disease (SCID),  
 CC adult respiratory distress syndrome, allergies, amyloidosis,  
 CC anaemia, asthma, atherosclerosis, Crohn's disease, atopic  
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
 CC Graves' disease, multiple sclerosis, myasthenia gravis, myocardial or  
 CC pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,  
 CC psoriasis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome,  
 CC scleroderma, systemic sclerosis, trauma; neurological disorders include  
 CC Alzheimer's disease, Huntington's disease, dementia, epilepsy,  
 CC Parkinson's disease, mental retardation and other developmental  
 CC disorders of central nervous system such as Down's syndrome, cerebral  
 CC palsy, periodic paralysis, mental disorders including mood, anxiety,  
 CC and schizophrenic disorders, seasonal affective disorder such as  
 CC akathesia, amnesia, cataplexy, dyskinesia; developmental disorders  
 CC include e.g. renal tubular acidosis, Duchenne and Becker muscular  
 CC dystrophy, gonadal dysgenesis, hypothyroidism; cell proliferative  
 CC disorders include e.g. actinic keratosis, arteriosclerosis,  
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis, and  
 CC cancer including adenocarcinoma, leukaemia. The polypeptide and  
 CC polynucleotide are further useful for analysing proteins of a tissue  
 CC or a cell type, for screening an agonist/antagonist, a compound that  
 CC specifically binds to it or its modulator. The polynucleotide is useful  
 CC for creating knockin humanised animals (pigs) or transgenic animals  
 CC (mice or rats) to model human disease, for generating a transcript image  
 CC of a tissue or cell type, which represents the global pattern of gene  
 CC expression by a particular tissue or cell type.  
 CC  
 XX Sequence 420 AA:  
 SO

Query Match 50.9%; Score 188; DB 23; Length 420;  
 Best Local Similarity 97.2%; Pred. No. 1.8e-18;  
 Matches 35; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MOEFMALELNLPGEFNSGNPSNRKRRYDILP 36  
 DB 160 MOEFMALELNLPGEFNSGNPSNRKRRYDILP 195

## RESULT 3

ABG06042

ID ABG06042 standard; Protein; 561 AA.

AC ABG06042;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #6033.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX MO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HXSE-) HTSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX MPI; 2001-639362/73.

XX N-PSDB; AAS70229.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT biologically for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 20; SEQ ID NO 36401; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridization probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 561 AA;

Query Match 48.8%; Score 180; DB 22; Length 561;

Best Local Similarity 94.4%; Pred. No. 3.8e-17;

Matches 34; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MOEFMALEIKNLPGFENSGNOPSNRKRNRYRDIILPF 36

DB 301 MOEFMALEIKNLPGFENSGNOPSNRKRNRYRDIILPF 336

RESULT 4

AAM89251

ID AAM89251 standard; Protein; 405 AA.

XX AAM89251;

XX 10-MAR-1999 (first entry)

DE Mouse PTP05 isoform #2.

KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease.

XX Mus sp.

XX MO9849317-A2.

XX 05-NOV-1998.

XX 27-APR-1998; 98WO-US08439.

XX 23-OCT-1997; 97US-0063595.

XX 28-APR-1997; 97US-0044428.

XX 20-MAY-1997; 97US-0047222.

XX 11-JUN-1997; 97US-0049477.

XX 11-JUN-1997; 97US-0049756.

XX 18-JUN-1997; 97US-0049914.

XX (SUGC-) SUGEN INC.

XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;

XX Marky D, Onrust S, Peles E, Plozman GD;

XX MPI; 1999-009434/01.

XX N-PSDB; AAV81746.

PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease

XX Claim 2; Page 158-160; 193pp; English.

XX The present invention describes isolated, enriched or purified nucleic  
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
XX present sequence represents mouse PTP05. The above proteins, other than  
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
XX substances that modulate their activity (i.e. agonists and antagonists,  
XX including NBP) in vivo or in vitro. These substances are used to treat  
XX or prevent diseases associated with abnormal signal transduction  
XX pathways that involve the proteins, particularly cancer (e.g. leukaemia  
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor  
XX serine/threonine kinase) are used to promote neuronal survival.  
XX Particularly for treating Alzheimer's, Parkinson's or Huntington's  
XX diseases. Nucleic acid fragments of the polynucleotides encoding the  
XX proteins can be used as probes to identify and clone related sequences;  
XX to detect protein-encoded RNA; to generate transgenic animals and in  
XX gene therapy (optionally after mutation). Ab are used to determine the  
XX proteins.

SQ Sequence 405 AA;

Query Match 32.8%; Score 121; DB 20; Length 405;

Best Local Similarity 61.1%; Pred. No. 1e-08;

Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MOEFMALEIKNLPGFENSGNOPSNRKRNRYRDIILPF 36

DB 166 IREFLEQMTLPDDFNSGNTLONDRDKNRYRDIILPF 201

RESULT 5

AAM89249

ID AAM89249 standard; Protein; 426 AA.

AC AAW89249;  
XX 10-MAR-1999 (first entry)  
XX  
XX  
XX Mouse PTP05.  
XX  
XX PPP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
XX type I receptor serine/threonine kinase; cancer; leukemia; lymphoma;  
XX neurodegenerative disease; neuronal survival; Alzheimer's disease;  
XX Parkinson's disease; Huntington's disease.  
XX  
XX Mus sp.  
XX  
XX WO9849317-A2.  
XX  
XX 05-NOV-1998.  
XX  
XX 27-APR-1998; 98WO-US08439.  
XX  
XX 23-OCT-1997; 97US-0063595.  
XX 28-APR-1997; 97US-0044428.  
XX 20-MAY-1997; 97US-0047222.  
XX 11-JUN-1997; 97US-0049477.  
XX 11-JUN-1997; 97US-0049756.  
XX 18-JUN-1997; 97US-0049914.  
XX  
XX (SUGEN-) SUGEN INC.  
XX  
XX App H, Clary D, Courtneldge SA, Hui TH, Jallal B;  
XX Markby D, Onrust S, Peles E, Plovman GD;  
XX  
XX WPI: 1999-009434/01.  
XX N-PSDB: AAV81744.  
XX  
XX New nucleic acid encoding specific protein tyrosine phosphatases -  
XX useful for identifying specific modulators for treatment and  
XX prevention of cancer and neurodegenerative disease  
XX  
XX  
XX Claim 2; Page 155-157; 193pp; English.  
XX  
XX The present invention describes isolated, enriched or purified nucleic  
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
XX present sequence represents mouse PTP05. The above proteins, other than  
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
XX substances that modulate their activity (i.e. agonists and antagonists,  
XX including NBP) in vivo or in vitro. These substances are used to treat  
XX or prevent diseases associated with abnormal signal transduction  
XX pathways that involve the proteins, particularly cancer (e.g. leukemia  
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor  
XX serine/threonine kinase) are used to promote neuronal survival,  
XX particularly for treating Alzheimer's, Parkinson's or Huntington's  
XX diseases. Nucleic acid fragments of the polynucleotides encoding the  
XX proteins can be used as probes to identify and clone related sequences;  
XX to detect protein-encoded RNA; to generate transgenic animals and in  
XX gene therapy (optionally after mutation). Ab are used to determine the  
XX proteins.  
XX  
XX  
XX Sequence 426 AA;  
XX  
XX  
XX Query Match 32.8%; Score 121; DB 20; Length 426;  
XX Best Local Similarity 61.1%; Pred. No. 1.1e-08;  
XX Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
XX  
XX 1 MOEFMALELNKLPGEFNSGNOPSNREKRYRDLPLP 36  
XX ::::: || ||:||||| ||:|||||  
XX Db 166 IREFLEQMTLPDDFNSGNTLQNRDKRYRDLPLP 201

RESULT 6  
AAW89250  
ID AAW89250 standard; Protein; 463 AA.  
XX

AC AAW89250;  
XX 10-MAR-1999 (first entry)  
XX  
XX  
XX Mouse PTP05 isoform #1.  
XX  
XX PPP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
XX type I receptor serine/threonine kinase; cancer; leukemia; lymphoma;  
XX neurodegenerative disease; neuronal survival; Alzheimer's disease;  
XX Parkinson's disease; Huntington's disease.  
XX  
XX Mus sp.  
XX  
XX WO9849317-A2.  
XX  
XX 05-NOV-1998.  
XX  
XX 27-APR-1998; 98WO-US08439.  
XX  
XX 23-OCT-1997; 97US-0063595.  
XX 28-APR-1997; 97US-0044428.  
XX 20-MAY-1997; 97US-0047222.  
XX 11-JUN-1997; 97US-0049477.  
XX 11-JUN-1997; 97US-0049756.  
XX 18-JUN-1997; 97US-0049914.  
XX  
XX (SUGEN-) SUGEN INC.  
XX  
XX App H, Clary D, Courtneldge SA, Hui TH, Jallal B;  
XX Markby D, Onrust S, Peles E, Plovman GD;  
XX  
XX WPI: 1999-009434/01.  
XX N-PSDB: AAV81745.  
XX  
XX New nucleic acid encoding specific protein tyrosine phosphatases -  
XX useful for identifying specific modulators for treatment and  
XX prevention of cancer and neurodegenerative disease  
XX  
XX  
XX Claim 2; Page 157-158; 193pp; English.  
XX  
XX The present invention describes isolated, enriched or purified nucleic  
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
XX present sequence represents mouse PTP05. The above proteins, other than  
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
XX substances that modulate their activity (i.e. agonists and antagonists,  
XX including NBP) in vivo or in vitro. These substances are used to treat  
XX or prevent diseases associated with abnormal signal transduction  
XX pathways that involve the proteins, particularly cancer (e.g. leukemia  
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor  
XX serine/threonine kinase) are used to promote neuronal survival,  
XX particularly for treating Alzheimer's, Parkinson's or Huntington's  
XX diseases. Nucleic acid fragments of the polynucleotides encoding the  
XX proteins can be used as probes to identify and clone related sequences;  
XX to detect protein-encoded RNA; to generate transgenic animals and in  
XX gene therapy (optionally after mutation). Ab are used to determine the  
XX proteins.  
XX  
XX  
XX Sequence 463 AA;  
XX  
XX  
XX Query Match 32.8%; Score 121; DB 20; Length 463;  
XX Best Local Similarity 61.1%; Pred. No. 1.2e-08;  
XX Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
XX  
XX 1 MOEFMALELNKLPGEFNSGNOPSNREKRYRDLPLP 36  
XX ::::: || ||:||||| ||:|||||  
XX Db 203 IREFLEQMTLPDDFNSGNTLQNRDKRYRDLPLP 238

RESULT 7  
AAW27226  
ID AAW27226 standard; Protein; 699 AA.  
XX  
XX AAW27226;  
XX

XX 19-DEC-1997 (first entry)  
 XX Mouse protein tyrosine phosphatase PTPepsilon.  
 XX Protein tyrosine phosphatase PTP-OB; PTPepsilon; osteoblast;  
 XX recombinant protein; growth; differentiation; brain; mouse.  
 XX Mus musculus.  
 XX Key Location/Qualifiers  
 XX Peptide 1..17  
 XX /label- Signal  
 XX Region 48..67  
 XX /label- Transmembrane  
 XX Region 154..400  
 XX /note- "Conserved PTP region"  
 XX Region 452..699  
 XX /note- "Conserved PTP region"  
 XX US5658756-A.  
 XX 19-AUG-1997.  
 XX 14-SEP-1993; 93US-0122032.  
 XX 01-DEC-1994; 94US-0348006.  
 XX 14-SEP-1993; 93US-0122032.  
 XX (MERI ) MERCK & CO INC.  
 XX Rodan GA, Rutledge SJ, Schmidt A;  
 XX WPI; 1997-424232/39.  
 XX DNA encoding protein tyrosine phosphatase PTP-OB - Isolated from  
 XX human osteoblasts and useful for production of recombinant PTP-OB  
 XX PS Disclosure; Column 39-42; 34pp; English.  
 XX The present sequence represents mouse protein tyrosine phosphatase  
 XX (PTPepsilon) protein. The DNA encoding this protein is useful for the  
 XX production of the recombinant protein, which is a protein tyrosine  
 XX phosphatase which may be involved in the growth and differentiation  
 XX of osteoblasts and brain cells and is useful for identifying compounds  
 XX that modulate PTP-OB activity and as a therapeutic agent for treating  
 XX PTP-OB-related diseases.  
 XX Sequence 699 AA:  
 SQ  
 Query Match 24.0%; Score 88.5; DB 18; Length 699;  
 Best Local Similarity 28.9%; Pred. No. 0.0011;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
 QY 2 QEPALALKLNPGEFNSGNQPSNRKKNRYRDLIPFOH-----H 39  
 DB 136 EEFNSLPSGHIQGFELANKENREKKNRYRDLIPNDHCYVILSQVDGIPCSDYINASYID 195  
 QY 40 GY-----SGNERTT--FW 51  
 DB 196 GYKEKNKFTAAQGPKEQTVNDFW 218  
 RESULT 8  
 ID AAM94028 standard; Protein: 699 AA.  
 AC AAM94028;  
 DE 01-APR-1999 (first entry)  
 XX Mouse protein tyrosine phosphatase mPTPepsilon.  
 XX

KW Protein tyrosine phosphatase; PTP; PTP-OB; bone; brain; cancer;  
 KW osteoporosis.  
 XX Mus sp.  
 XX Key Location/Qualifiers  
 XX Peptide 1..17  
 XX /note- "signal peptide"  
 XX Region 48..66  
 XX /note- "transmembrane region"  
 XX Region 154..399  
 XX /note- "conserved PTP region"  
 XX Region 452..699  
 XX /note- "conserved PTP region"  
 XX US5866397-A.  
 XX 02-FEB-1999.  
 XX 14-FEB-1997; 97US-0800825.  
 XX 01-DEC-1994; 94US-0348006.  
 XX 14-SEP-1993; 93US-0122032.  
 XX 14-FEB-1997; 97US-0800825.  
 XX (MERI ) MERCK & CO INC.  
 XX Rodan GA, Rutledge SJ, Schmidt A;  
 XX WPI; 1999-141930/12.  
 XX Protein tyrosine phosphatase denoted PTP-OB - useful for drug  
 XX screening  
 XX PS Disclosure; Fig 11; 34pp; English.  
 XX The invention relates to a human protein tyrosine phosphatase (PTP)  
 XX denoted as PTP-OB, produced by bone and brain cells. A recombinant host  
 XX cell transfected or transformed with a nucleic acid vector comprising  
 XX the nucleic acid can be used for the production of the PTP-OB protein.  
 XX The protein can be used to screen for modulators of PTP-OB activity,  
 XX which might be useful for treating e.g. osteoporosis and cancer.  
 XX The present sequence represents a mouse mPTPepsilon polypeptide sequence.  
 XX Sequence 699 AA:  
 SQ  
 Query Match 24.0%; Score 88.5; DB 20; Length 699;  
 Best Local Similarity 28.9%; Pred. No. 0.0011;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
 QY 2 QEPALALKLNPGEFNSGNQPSNRKKNRYRDLIPFOH-----H 39  
 DB 136 EEFNSLPSGHIQGFELANKENREKKNRYRDLIPNDHCYVILSQVDGIPCSDYINASYID 195  
 QY 40 GY-----SGNERTT--FW 51  
 DB 196 GYKEKNKFTAAQGPKEQTVNDFW 218  
 RESULT 9  
 ID AAU01460 standard; Protein: 699 AA.  
 AC AAU01460;  
 DE 18-JUL-2001 (first entry)  
 XX Mouse protein tyrosine phosphatase, mPTP-epsilon, amino acid sequence.  
 XX Protein tyrosine phosphatase; mPTP-epsilon; mouse; osteoporosis;  
 KW bone formation; PTP-OB; cell proliferation; cell death; lung cancer;  
 KW neoplastic transformation; metastasis; tumour; cancer; osteosarcoma.  
 XX





XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder;  
 KW biodiversity; gene therapy; nutrition.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200154477-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 25-JAN-2001; 2001WO-US02687.  
 XX  
 PR 25-JAN-2000; 2000US-0491404.  
 PR 17-JUL-2000; 2000US-0617746.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 15-SEP-2000; 2000US-0663870.  
 XX  
 PA (HYSEQ-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
 XX  
 DR WPI: 2001-476164/51.  
 DR N-PSDB; AAH98486.  
 XX  
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX  
 PS Claim 20; Page 953-954; 1275pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 XX  
 SQ Sequence 700 AA;  
 Query Match 24.0%; Score 88.5; DB 22; Length 700;  
 Best Local Similarity 28.9%; Pred. No. 0.0011;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
 QY 2 OEFALELKLNPGEFNSGNQPSNRKRYRDLPPQH-----H 39  
 DB 137 EEFNLSLPSGHIGTFELANKRENRKRYRDLPPDHRSRVLSQLDGIPCSDIYINASTYD 196  
 QY 40 GY-----SGPNERTT--FW 51  
 DB 197 GYKEKNKFIAAGPKQETVNDFW 219  
 XX  
 RESULT 12  
 ID AAM38831 standard; Protein: 700 AA.  
 AC AAM38831;  
 XX  
 XX 22-OCT-2001 (first entry)  
 DE Human polypeptide SEQ ID NO 1976.  
 XX  
 KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW Leukaemia.  
 XX

OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0633450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSEQ-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AAI57987.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 3; SEQ ID NO 1976; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AAI5798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AA42213) with noctropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localized neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 700 AA;  
 Query Match 24.0%; Score 88.5; DB 22; Length 700;  
 Best Local Similarity 28.9%; Pred. No. 0.0011;  
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;  
 QY 2 OEFALELKLNPGEFNSGNQPSNRKRYRDLPPQH-----H 39  
 DB 137 EEFNLSLPSGHIGTFELANKRENRKRYRDLPPDHRSRVLSQLDGIPCSDIYINASTYD 196  
 QY 40 GY-----SGPNERTT--FW 51  
 DB 197 GYKEKNKFIAAGPKQETVNDFW 219  
 XX  
 RESULT 13  
 ID AAM40617 standard; Protein: 717 AA.  
 AC AAM40617;  
 XX  
 XX 22-OCT-2001 (first entry)  
 DE Human polypeptide SEQ ID NO 5548.  
 XX

KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX Homo sapiens.

XX MO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-063450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB: AAI59773.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -

XX Example 2; SEQ ID NO 5548; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AA138642-AA142213) with noctropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX Sequence 717 AA;

XX Query Match 24.0%; Score 88.5; DB 22; Length 717;

XX Best Local Similarity 28.9%; Pred. No. 0.0012;

XX Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

XX 2 QEFMALTEKNLPGFENSGNOPSNRKKNRYRDLIPFOH-----H 39

XX 154 EEFNSLPGHIOGTFLANKNEKKNRPNTLPNDHSVILLSQDGPICSDYINASYID 213

XX 40 GY-----SGPNERTT--FW 51

XX 214 GYKKNKFIAMQGPKEFTVNDWF 236

XX RESULT 14  
 XX AAR99314  
 XX ID AAR99314 standard; Protein; 513 AA.

XX AAR99314;

XX 13-NOV-1996 (first entry)

XX Rat M1PRP (protein tyrosine phosphatase-1) clone.

XX PTP; protein tyrosine phosphatase; SH2; Src homology region 2;

XX chromosome 12p; abnormality; mutation; detection; probe; neoplasia;  
 KW cancer; leukaemia; diagnosis; megakaryocyte regulation.

XX Rattus sp.

XX US5536636-A.

XX 16-JUL-1996.

XX 26-JUN-1991; 91US-0721112.

XX 28-FEB-1994; 94US-0202389.

XX 26-JUN-1991; 91US-0721112.

XX 31-JAN-1992; 92US-0829141.

XX 01-DEC-1992; 92US-0983926.

XX (BETH-) BETH ISRAEL HOSPITAL ASSOC.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Freeman RM, Neel BG, Plutsky J, Rosenberg RD;

XX WPI: 1996-341506/34.

XX N-PSDB: AAR35308.

XX Detecting 12p chromosomal abnormality associated with neoplastic  
 PT disease - using SH-PRP1 protein tyrosine phosphatase gene specific  
 PT probe

XX Example 1; Column 32-34; 63pp; English.

XX AAR99314 is a rat M1PRP (protein tyrosine phosphatase-1) clone. A cDNA  
 CC clone encoding this sequence was used as a probe to isolate a  
 CC series of overlapping human cDNA clones defining the human M1PRP  
 CC equivalent, SH-PRP1, which contains two tandem SH2 domains. A  
 CC sequence complementary to nucleotides 537-653 of the SH-PRP1 gene  
 CC clone given as AAR35306 is used as a probe to detect a chromosome  
 CC 12p13 abnormality associated with neoplastic disease, in partic.  
 CC acute lymphoblastic leukaemia. The probe hybridises to a part  
 CC of the region coding for the two tandem SH2 domains (see AAR99312).  
 CC If the probe will not hybridise DNA of chromosome 12p13 from a patient  
 CC sample it is indicative of an abnormality, normally associated with  
 CC neoplasia. Fragments of sequences encoding human SH-PRP2 (see AAR35307)  
 CC may also be used to diagnose a condition or susceptibility to a  
 CC condition associated with chromosome 12p abnormality. Alternatively  
 CC the wild-type SH-PRP1 or SH-PRP2 gene or protein may be used for  
 CC comparison to sequenced PTP genes taken from a patient, where  
 CC differences indicate an abnormality. The activity of SH-PRP1 may also  
 CC be purposely altered by mutation to effect a change in megakaryocyte  
 CC function and hence platelet production.

XX Sequence 513 AA;

XX Query Match 23.8%; Score 88; DB 17; Length 513;

XX Best Local Similarity 47.5%; Pred. No. 0.00089;

XX Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

XX 2 QEFMAL---ELKNLPGFENSGNOPSNRKKNRYRDLIPFOH 38

XX 164 EEFESLQKQEVKNLHORL-EGORPENKSKKNRYNLPFDH 202

XX RESULT 15  
 XX AAR99316  
 XX ID AAR99316 standard; Protein; 458 AA.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 11, 2003, 12:55:32 ; Search time 64 Seconds

(Without alignments)  
321,052 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369  
Sequence: 1 MOEFMALEKMLPGEFNSGN.....TTFMHSNEGAVSLIRYCA 67

Scoring table:

BLOSUM62  
Xgapop 10.0, Ygapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL-frame-p2n.model -DEV-xlp  
-O=/cgn2\_1/USPRO.spool/US09935703/runat\_06062003\_095212\_6181/app-query.fasta.1.263  
-DB-issued\_patents.NA -QEMT-fastcap -SUFFIX-p2n.rml -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cd1  
-LIST=45 -DOCALLIGN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE-LOCAL -OUTFMT=pro -NORM-ext HEMASIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09935703\_ECGN\_1.1.61.etruncat\_06062003\_095212\_6181 -NCP=6 -ICPU=3  
-NO\_MMAP -LARGESOURCY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database:

1: Issued\_Patents.NA:\*  
2: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/BACKFILES1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	23.8	1747	1	US-08-202-389-1
2	88	23.3	2143	1	US-08-202-389-7
3	86	23.3	2145	1	US-08-202-388-5
4	86	23.3	2277	3	US-09-358-685-1
5	79.5	21.5	2409	1	US-08-015-985-2
6	79.5	21.5	2872	1	US-08-015-985-4
7	78.5	21.3	466	4	US-08-821-2278A-16
8	73.5	19.9	1659	4	US-09-152-060-24
9	73	19.8	878	2	US-08-446-345-33
10	73	19.8	4080	2	US-08-446-345-35
11	72	19.5	2121	4	US-09-358-683-1
12	72	19.5	2276	1	US-08-202-389-11

13	71	19.2	247	1	US-08-594-031-74	Sequence 74, Appl
14	70	19.0	612	4	US-09-280-116-218	Sequence 218, App
15	70	19.0	2790	2	US-08-018-129-4	Sequence 4, Appl1
16	70	19.0	2790	2	US-08-448-250-4	Sequence 4, Appl1
17	68	18.4	963	3	US-08-964-308-1	Sequence 1, Appl1
18	68	18.4	963	3	US-08-964-308-7	Sequence 7, Appl1
19	68	18.4	963	3	US-08-964-313-1	Sequence 1, Appl1
20	68	18.4	963	3	US-08-964-313-7	Sequence 7, Appl1
21	68	18.4	963	4	US-09-063-138-1	Sequence 1, Appl1
22	68	18.4	963	4	US-09-063-138-7	Sequence 7, Appl1
23	68	18.4	3247	4	US-09-487-368A-3	Sequence 3, Appl1
24	68	18.4	3969	1	US-08-241-853-1	Sequence 1, Appl1
25	68	18.4	4369	2	US-08-850-917-1	Sequence 1, Appl1
26	68	18.4	4127	4	US-09-487-368A-10	Sequence 10, Appl1
27	67.5	18.3	2226	3	US-08-951-260A-1	Sequence 1, Appl1
28	67.5	18.3	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
29	67.5	18.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
30	66.5	18.0	4752	1	US-08-201-697-3	Sequence 3, Appl1
31	66.5	18.0	4815	1	US-08-201-697-5	Sequence 5, Appl1
32	66.5	18.0	5679	1	US-08-201-697-1	Sequence 1, Appl1
33	65.5	17.8	4799	1	US-08-201-697-6	Sequence 6, Appl1
34	64.5	17.5	3580	4	US-09-081-345-1	Sequence 1, Appl1
35	63	17.1	1079	4	US-09-221-017B-540	Sequence 540, App
36	62.5	16.9	2577	2	US-08-439-818A-4	Sequence 4, Appl1
37	62.5	16.9	2577	2	US-08-751-965-4	Sequence 4, Appl1
38	62.5	16.9	2577	2	US-08-738-975-4	Sequence 4, Appl1
39	62.5	16.9	2577	2	US-08-728-626-4	Sequence 4, Appl1
40	62.5	16.9	2577	3	US-08-808-599A-4	Sequence 4, Appl1
41	62.5	16.9	2578	1	US-08-317-522A-4	Sequence 4, Appl1
42	62	16.8	610	4	US-09-385-982-229	Sequence 229, App
43	62	16.8	1050	4	US-08-971-090-4	Sequence 4, Appl1
44	62	16.8	6924	1	US-08-015-973-2	Sequence 2, Appl1
45	62	16.8	6924	2	US-08-448-164-2	Sequence 2, Appl1

RESULT 1  
US-08-202-389-1  
Sequence 1, Application US/08202389  
Patent No. 5536636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutzky, Jorge  
APPLICANT: Neel, Benjamin G.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112

FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIH92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1747 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..1540  
US-08-202-389-1

Alignment Scores:  
Pred. No.: 0.00235 Length: 1747  
Score: 88.00 Matches: 19  
Percent Similarity: 65.00% Conservative: 7  
Best Local Similarity: 47.50% Mismatches: 10  
Query Match: 23.85% Indels: 4  
Gaps: 2

US-09-935-703-7 (1-67) x US-08-202-389-1 (1-1747)

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18  
Db 491 GAGGAGTTTGAGAGTGTGCAAAAGCAAGAGGTAAAGACTTGACACCGCTCG---GAA 547  
QY 19 GlysGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38  
Db 548 GGGCAGCGGGCGGAGAGCAAGACGACGAGCAAGACCGCTACAGAACATCTCCCTTTGACAC 607

RESULT 2  
US-08-202-389-7  
Sequence 7, Application US/08202389  
Patent No. 5536636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutsky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Rosenberg, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIH92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2143 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 145..2037  
US-08-202-389-7

Alignment Scores:  
Pred. No.: 0.00624 Length: 2143  
Score: 86.00 Matches: 19  
Percent Similarity: 65.00% Conservative: 7  
Best Local Similarity: 47.50% Mismatches: 10  
Query Match: 23.31% Indels: 4  
Gaps: 2

US-09-935-703-7 (1-67) x US-08-202-389-7 (1-2143)

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18  
Db 880 GAGGAGTTTGAGAGTGTGCAAGAGCAAGAGGTAAAGACTTGACACCGCTCG---GAA 936  
QY 19 GlysGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38  
Db 937 GGGCAGCGGGCGGAGAGCAAGAGGCAAGACCGCTACAGAACATCTCCCTTTGACAC 996

RESULT 3  
US-08-202-389-5  
Sequence 5, Application US/08202389  
Patent No. 5536636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutsky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Rosenberg, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: B1H92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2145 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 145..1929  
US-08-202-389-5

Alignment Scores:  
Pred. No.: 0.00624 Length: 2145  
Score: 86.00 Matches: 19  
Percent Similarity: 65.00% Conservative: 7  
Best Local Similarity: 47.50% Mismatches: 10  
Query Match: 23.31% Indels: 4  
DB: 1 Gaps: 2

US-09-935-703-7 (1-67) x US-08-202-389-5 (1-2145)

Qy 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

Db 880 GAGGAGTTTGAGAGTTGCAGACGAGGAGGAGAACTTGACCGCTCTG--GAA 936

Qy 19 GlysAnglnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38  
Db 937 GGGCAGCGCCAGACAGAACAGGCGCAAGAACCCCTACAAACATTTCTCCCTTTGACAC 996

## RESULT 4

US-09-358-685-1  
Sequence 1, Application US/09358685  
Patent No. 6121047  
GENERAL INFORMATION:  
APPLICANT: C. Frank Bennett  
TITLE OF INVENTION: ANTISENSE MODULATION OF SHP-1 EXPRESSION  
FILE REFERENCE: RTS-0081  
CURRENT APPLICATION NUMBER: US/09/358,685  
CURRENT FILING DATE: 1999-07-21  
NUMBER OF SEQ ID NOS: 47  
SEQ ID NO 1  
LENGTH: 2277  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (256)..(2049)  
US-09-358-685-1

Alignment Scores:  
Pred. No.: 0.0068 Length: 2277  
Score: 86.00 Matches: 19  
Percent Similarity: 65.00% Conservative: 7  
Best Local Similarity: 47.50% Mismatches: 10  
Query Match: 23.31% Indels: 4  
DB: 3 Gaps: 2

US-09-935-703-7 (1-67) x US-09-358-685-1 (1-2277)

Qy 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

Db 997 GAGGAGTTTGAGACTTTCAGACAGGAGGAGTGAAGAACTTGACACCGCTCTG--GAA 1053

Qy 19 GlysAnglnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38  
Db 1054 GGGCAGCGCCAGACAGAACAGGCGCAAGAACCCCTACAAACATTTCTCCCTTTGACAC 1113

## RESULT 5

US-08-015-985-2  
Sequence 2, Application US/08015985  
Patent No. 5538886  
GENERAL INFORMATION:  
APPLICANT: Schlessinger, Joseph  
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,985  
FILING DATE: 10-FEB-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7683-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2409 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-015-985-2

Alignment Scores:  
Pred. No.: 0.0681 Length: 2409  
Score: 79.50 Matches: 22  
Percent Similarity: 37.35% Conservative: 9  
Best Local Similarity: 26.51% Mismatches: 19  
Query Match: 21.54% Indels: 33  
DB: 1 Gaps: 3

US-09-935-703-7 (1-67) x US-08-015-985-2 (1-2409)

Qy 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21

Db 727 GAGGAATTTCACAGCTTCCTCCGATGTCCTATCCAGGCCACCTGTGAGGCTCTCCCAAG 786

Qy 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis----- 38

Db 787 GAGGAAAACAAGGAAAAAATCGATATCTAACAATCTTCCTTATGACCACTCTAGAGTC 846

Qy 39 -----His 39

Db 847 CACCTGACACCGGTGAGGAGGCTTCAGATTCTGATTACATCAATGCTTCAATCAAC 906

Qy 40 GlyTyr-----SerglyProAsnGluArgThrThr--- 49

Db	967	GATTCCAGAAAGACAATTCATTGCTGCACAAAGACCAAAAGAGAAACGGTCAT	966
Qy	50	---PheTrp	51
Db	967	GATTTCTG	975

Qy	50	--Phe	trp	51
Db	1138	GGCTACCAAGAAACAAATTCATCGCTGCACAGGACCAAAAAGAAACACTGAT	1197	
Db	1198	GACTTCTCG	1206	

```

1  APPLICATION NUMBER: US 08/234,440
2  FILING DATE: 28-APR-1994
3  ATTORNEY/AGENT INFORMATION:
4  NAME: CORUZZI, Laura A.
5  REGISTRATION NUMBER: 30742
6  REFERENCE/DOCKET NUMBER: 7663-054
7  TELECOMMUNICATION INFORMATION:
8  TELEPHONE: (212)790-9090
9  TELEFAX: (212) 869-8864
10  TELEX: 66141 PENNTE
11  INFORMATION FOR SEQ ID NO: 33:
12  SEQUENCE CHARACTERISTICS:
13  LENGTH: 878 base pairs
14  TYPE: nucleic acid
15  STRANDEDNESS: both
16  TOPOLOGY: unknown
17  MOLECULE TYPE: cDNA
18  US-08-446-345-33
19
20  Alignment Scores:
21  Pred. No.: 0.148 Length: 878
22  Score: 73.00 Matches: 17
23  Percent Similarity: 59.57% Conservative: 11
24  Best Local Similarity: 36.17% Mismatches: 17
25  Query Match: 19.78% Indels: 2
26  DB: 2 Gaps: 0
27
28  US-09-935-703-7 (1-67) x US-08-446-345-33 (1-878)
29
30  QY 7 LeuGIuLeuLYsAsnLeuProGIyGIuPheAsnSerGIyAsnGlnProSerAspArgGlu 26
31  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
32  Db 6 CTTAAGAAAGCGCTGATGATGGGAGGTCTCAACAGCAGCACTCCCTGAAATGCAGAA 65
33
34  QY 27 LysAsnArgTYArgAspIleLeuProPheGlnHis-HisGIYTYSer-GlyProAsnG 46
35  ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
36  Db 66 AGAATCGATGCCAAGATGTCCTCTCTATGATGATGCAGATGGAGTGGATGCCCACT 125
37  QY 46 LuArgThrThrPheTrp 51
38  ||| ||| |||
39  Db 126 AAAAGAAACMACACTGG 142
40
41  RESULT 10
42  US-08-446-345-35
43  Sequence 35, Application US/08446345
44  Patent No. 5831009
45  GENERAL INFORMATION:
46  APPLICANT: Ullrich, Axel
47  APPLICANT: Moller, Niels P.H.
48  APPLICANT: Moller, Karin B.
49  TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
50  TITLE OF INVENTION: PHOSPHATASES PTP-DI
51  NUMBER OF SEQUENCES: 41
52  CORRESPONDENCE ADDRESS:
53  ADDRESSEE: Pennie & Edmonds
54  STREET: 1155 Avenue of the Americas
55  CITY: New York
56  STATE: N.Y.
57  COUNTRY: U.S.A.
58  ZIP: 10036-2711
59  COMPUTER READABLE FORM:
60  MEDIUM TYPE: Floppy disk
61  COMPUTER: IBM PC compatible
62  OPERATING SYSTEM: PC-DOS/MS-DOS
63  SOFTWARE: PatentIn Release #1.0, Version #1.25
64  CURRENT APPLICATION DATA:
65  APPLICATION NUMBER: US/08/446,345
66  FILING DATE: 22-MAY-1995
67  CLASSIFICATION: 435
68  PRIOR APPLICATION DATA:
69  APPLICATION NUMBER: US 08/234,440
70  FILING DATE: 28-APR-1994
71  ATTORNEY/AGENT INFORMATION:
72  NAME: CORUZZI, Laura A.

```



REGISTRATION NUMBER: 30742  
REFERENCE/DOCKET NUMBER: 7683-054  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4080 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
US-08-446-345-35

Alignment Scores:  
Pred. No.: 1.34 Length: 4080  
Score: 73.00 Matches: 17  
Percent Similarity: 59.57% Conservative: 11  
Best Local Similarity: 36.17% Mismatches: 17  
Query Match: 19.78% Indels: 2  
DB: 2 Gaps: 0

US-09-935-703-7 (1-67) x US-08-446-345-35 (1-4080)

OY 7 LeuGlutLeuLysAsnLeuProGlyLupheAsnSerGlyAsnGlnProSerAsnArgIu 26  
Db 3042 CTTAAGAAAGCGGCGATTGATGGGAGTGCCTCAACAGCAGCAGCTCCCGAAATGCAGAA 3101

OY 27 LysAsnArgTyrArgAspIleLeuProPheGlnHis-HisGlyTyrSer-GlyProAsn 46  
Db 3102 AGAAATCGATTCACAGATGTTCTTCCTTATGATGCGAGAGTGGAGTGTGCTCCCAACT 3161

OY 46 LuArgThrThrPheTrp 51  
Db 3162 AAAAGAAACACACTGG 3178

RESULT 11  
US-09-358-683-1  
Sequence 1, Application US/09358683  
Patent No. 6200807  
GENERAL INFORMATION:  
APPLICANT: C. Frank Bennett  
TITLE OF INVENTION: ANTISENSE MODULATION OF SHP-2 EXPRESSION  
FILE REFERENCE: RTS-0082  
CURRENT APPLICATION NUMBER: US/09/358,683  
CURRENT FILING DATE: 1999-07-21  
NUMBER OF SEQ ID NOS: 47  
SEQ ID NO 1  
LENGTH: 2121  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (154)..(1935)  
US-09-358-683-1

Alignment Scores:  
Pred. No.: 0.737 Length: 2121  
Score: 72.00 Matches: 16  
Percent Similarity: 54.76% Conservative: 7  
Best Local Similarity: 38.10% Mismatches: 13  
Query Match: 19.51% Indels: 6  
DB: 4 Gaps: 2

US-09-935-703-7 (1-67) x US-09-358-683-1 (1-2121)

OY 19 GlyAsnGlnProSerAsnArgGlyLupheAsnArgTyrArgAspIleLeuProPheGlnHis 38  
Db 955 GGTCAAGGCAAGCAAAACAAACAAATAGATTAATAAAATCCTGCTTGTGATCAT 1014

OY 39 -----HisGlyTyrSerGlyProAsnGluArgThrThrPheTrpHisGly 53

Db 1015 ACCAGGCTGCTCCTACAGAT---GGTATCCCAATGAGCCTGTTTCAGATTACATCAAT 1071  
OY 54 SerAsn 55  
Db 1072 GCAAT 1077

RESULT 12  
US-08-202-389-11

Sequence 11, Application US/08202389  
Patent No. 553636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutsky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Rosenberg, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIR92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2276 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 114..1893  
US-08-202-389-11

Alignment Scores:  
Pred. No.: 0.815 Length: 2276  
Score: 72.00 Matches: 16  
Percent Similarity: 54.76% Conservative: 7  
Best Local Similarity: 38.10% Mismatches: 13  
Query Match: 19.51% Indels: 6  
DB: 1 Gaps: 2

US-09-935-703-7 (1-67) x US-08-202-389-11 (1-2276)



CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/018.129  
FILING DATE: 19930216  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2790 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 130..1911  
US-08-018-129-4

Alignment Scores:  
Pred. No.: 2.16  
Score: 70.00  
Percent Similarity: 53.66%  
Best Local Similarity: 36.59%  
Query Match: 18.97%  
DB: 1  
Length: 2790  
Matches: 15  
Conservative: 7  
Mismatch: 15  
Indels: 4  
Gaps: 1

US-09-935-703-7 (1-67) x US-08-018-129-4 (1-2790)

QY 19 GlyAsnGlnProSerAsnArgGluLyAsnArgTyrArgAspIleLeuProPheGlnHis 38  
DB 931 GGTCAAGGCGACAGAAACAAATAAGATATAAAACATCGCCCTTGATCAT 990  
QY 39 HIs-----GlyTyrSerGlyProAsnGluArgThrThrPheTrpHisGlySer 54  
DB 991 ACCAGGCTGTCTCTACGATCTGTGATCCCAATGAGCCGTTCAGATTACATCATGCA 1050  
QY 55 Asn 55  
DB 1051 AAT 1053

Search completed: June 11, 2003, 14:05:54  
Job time : 71 secs

Thu Jun 12 08:30:30 2003

us-09-935-703-7.p2n.rnpb

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame\_plus.p2n model

Run on: June 11, 2003, 13:40:28 ; Search time 117 Seconds

(Without alignments)  
801.444 Million cell updates/sec

Title: US-09-935-703-7  
Perfect score: 369  
Sequence: 1 MOEFMALEKLNLEPFNSGN.....TTFHNSGVALRYCA 67

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 870385 seqs, 69976693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ p2n.model -DEV=rip  
-O/cgcn2.1/USPTO.spool/US0935703/rnucat\_06062003-095213-6265/app.query.fasta.1.263  
-DB-published.Applications.NA -OFMT-fasta -SUFFIX-p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62  
-TRANS-human40.csl -LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100  
-THR.MIN=0 -ALIGN=15 -MODE-LOCAL -OUTFMT-pto -NORM-ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US0935703.ecgn.1.1.80 -rnucat\_06062003-095213-6265  
-NCP=6 -ICPU=3 -NO.MMAP -LARGQUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FAPOP=6 -FAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published.Applications.NA.\*

- 1: /cgcn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgcn2.6/ptodata/1/pubpna/PTCT\_NEW\_PUB.seq.\*
- 3: /cgcn2.6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgcn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 5: /cgcn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 6: /cgcn2.6/ptodata/1/pubpna/PTCT\_NEW\_PUB.seq.\*
- 7: /cgcn2.6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgcn2.6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgcn2.6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgcn2.6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgcn2.6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgcn2.6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgcn2.6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgcn2.6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.5	24.0	2389	9	US-10-098-841-192 Sequence 192, App
2	88	23.8	402	10	US-09-960-352-3199 Sequence 3635, Ap
3	86	23.3	326	9	US-09-796-692-3635 Sequence 5490, Ap
4	86	23.3	326	9	US-09-796-692-3635 Sequence 5490, Ap

5	86	23.3	326	9	US-10-040-862-3635	Seq.
6	86	23.3	326	9	US-10-040-862-3635	Sequence 3721, Ap
7	86	23.3	333	9	US-09-796-692-3721	Sequence 5, App1
8	86	23.3	333	9	US-10-040-862-3721	Sequence 2, App1
9	86	23.3	2076	10	US-09-822-849A-5	Sequence 1478, Ap
10	86	23.3	2145	10	US-09-920-021A-2	Sequence 118, App
11	86	23.3	2176	10	US-09-917-800A-1478	Sequence 118, App
12	86	23.3	2532	9	US-10-072-036-116	Sequence 31823, A
13	86	23.3	2562	9	US-10-072-036-118	Sequence 5639, Ap
14	86	23.3	3220	12	US-10-002-600-96	Sequence 30413, A
15	81.5	22.1	535	9	US-09-918-985-1823	Sequence 35, App1
16	81.5	22.1	14738	9	US-09-764-891-5637	Sequence 11796, A
17	81.5	22.1	14738	9	US-09-764-891-5637	Sequence 24, App1
18	79.5	21.5	564	9	US-09-764-891-5639	Sequence 24, App1
19	78.5	21.3	2810	9	US-09-918-995-30413	Sequence 342, App
20	73.5	19.9	488	10	US-09-864-761-11796	Sequence 21, App1
21	73.5	19.9	1659	9	US-09-852-797-24	Sequence 622, App
22	73.5	19.9	1659	10	US-09-853-161-24	Sequence 300, App
23	73.5	19.9	1659	10	US-09-852-659A-24	Sequence 1503, App
24	73	19.8	4080	10	US-09-962-436-342	Sequence 1375, Ap
25	72	19.5	1794	9	US-10-038-010-21	Sequence 1375, Ap
26	72	19.5	2276	10	US-09-920-021A-4	Sequence 1375, Ap
27	72	19.5	2300	12	US-10-044-090-622	Sequence 1375, Ap
28	70	19.0	587	10	US-09-925-300-300	Sequence 1375, Ap
29	70	19.0	1494	10	US-09-917-800A-1503	Sequence 1375, Ap
30	70	19.0	1551	9	US-10-180-921-1	Sequence 1375, Ap
31	68.5	18.6	295	9	US-09-736-457-1375	Sequence 1375, Ap
32	68.5	18.6	295	9	US-09-902-941-1375	Sequence 1375, Ap
33	68.5	18.6	295	9	US-09-849-626-1375	Sequence 1375, Ap
34	68.5	18.6	295	9	US-10-017-754-1375	Sequence 1375, Ap
35	68.5	18.6	3650	9	US-09-927-827-31	Sequence 241, App
36	68	18.4	2346	10	US-09-854-883-241	Sequence 1555, Ap
37	68	18.4	4127	10	US-09-854-883-3	Sequence 1555, Ap
38	68	18.4	4127	10	US-09-854-883-10	Sequence 1555, Ap
39	68	18.4	4127	10	US-09-917-800A-1555	Sequence 1555, Ap
40	67.5	18.3	446	10	US-09-867-701-7967	Sequence 1555, Ap
41	67.5	18.3	2226	9	US-10-087-993-31	Sequence 1555, Ap
42	67.5	18.3	2226	9	US-10-243-687-1	Sequence 1555, Ap
43	64.5	17.5	3580	10	US-09-822-295-1	Sequence 1555, Ap
44	64.5	17.5	3580	10	US-10-005-467-1	Sequence 1555, Ap
45	64.5	17.5	143306	10	US-09-729-920-3	Sequence 1555, Ap

ALIGNMENTS

RESULT 1  
US-10-098-841-192  
; Sequence 192, Application US/10098841  
; Publication No. US20020197679A1  
GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Xu, Chongjun  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungling  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: Weinman, Tom  
; APPLICANT: Zhang, Jie  
; APPLICANT: Qian, Xiaohong B.  
; APPLICANT: Drmanac, Radote T.  
TITLE OF INVENTION: NO. US20020197679A1 Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP2  
CURRENT APPLICATION NUMBER: US/10/098,841  
CURRENT FILING DATE: 2002-03-13  
PRIOR APPLICATION NUMBER: 09/598,042  
PRIOR FILING DATE: 2000-06-20

PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 331  
SOFTWARE: PL\_FL\_genes Version 1.0  
SEQ ID NO 192  
LENGTH: 2389  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (32)..(2154)  
US-10-098-841-192

Alignment Scores:  
Pred. No.: 0.00249

Score: 88.50  
Percent Similarity: 37.358  
Best Local Similarity: 28.928  
Query Match: 23.988  
Length: 2389  
Matches: 24  
Conservative: 7  
Mismatch: 19  
Indels: 33  
Gaps: 3

US-09-935-703-7 (1-67) x US-10-098-841-192 (1-2389)

QY 2 GINGLUPHEMETHALAEU-----GULEULYSASNLEUPROGLYGIUPHEANSERGLYASNGLN 21  
DB 460 GAGGAGTTGAGAGTGTGCGAAGACATACAGCAATTTGGAACGGCAATTA 519  
QY 22 PROSERASNAARGGLYASNAARGTYRARGASPLLEUPROPHGLNHIS----- 38  
DB 520 GAAGAAACAGCAAAACAGATATCCACATCTCCCAATGACCATTTAGGGTG 579  
QY 39 -----HIS 39  
DB 580 ATTCTGAGCCACTGATGGAATTCCTGTCAGCATCATGCTTCTCATATGAT 639  
QY 40 GITYR-----Serglyproasngluarqthr--- 49  
DB 640 GGTTCAAAGAGAGATTAATTCATAGCAGTCAAGTCCCAAGCAAGGTTAAC 699  
QY 50 ---PheTrp 51  
DB 700 GACTTCTGG 708

## RESULT 2

US-09-960-352-3199  
Sequence 3199, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16311.006/37-21(10288)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 3199  
LENGTH: 402  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 14-LIB34-017-Q1-E1-D9  
US-09-960-352-3199

## Alignment Scores:

Pred. No.: 0.000279  
Score: 88.00  
Percent Similarity: 65.008  
Best Local Similarity: 47.508  
Query Match: 23.858  
Length: 402  
Matches: 19  
Conservative: 7  
Mismatch: 10  
Indels: 4

DB: 10 Gaps: 2

US-09-935-703-7 (1-67) x US-09-960-352-3199 (1-402)

QY 2 GINGLUPHEMETHALAEU-----GULEULYSASNLEUPROGLYGIUPHEANSER 18  
DB 174 GAGGAGTTGAGAGTGTGCGAAGACATACAGCAATTTGGAACGGCAATTA 230  
QY 19 GLYASNGINPROSERASNAARGGLYASNAARGTYRARGASPLLEUPROPHGLNHIS 38  
DB 231 GGGCAGCGGCGCAAGAAACAGAGCAAGACCGTTACAGAACATCTCCGTTGACAC 290

## RESULT 3

US-09-796-692-3635  
Sequence 3635, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Galger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
PRIOR FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3635  
LENGTH: 326  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-3635

## Alignment Scores:

Pred. No.: 0.000424  
Score: 86.00  
Percent Similarity: 65.008  
Best Local Similarity: 47.508  
Query Match: 23.318  
Length: 326  
Matches: 19  
Conservative: 7  
Mismatch: 10  
Indels: 4  
Gaps: 2

US-09-935-703-7 (1-67) x US-09-796-692-3635 (1-326)

QY 2 GINGLUPHEMETHALAEU-----GULEULYSASNLEUPROGLYGIUPHEANSER 18  
DB 132 GAGGAGTTGAGAGTGTGCGAAGACATACAGCAATTTGGAACGGCAATTA 188  
QY 19 GLYASNGINPROSERASNAARGGLYASNAARGTYRARGASPLLEUPROPHGLNHIS 38

Thu Jun 12 08:30:30 2003

us-09-935-703-7.p2n.rnpb

DB 189 GGGCAGCGCCAGAGAACAGGCGACAGACCGCTACAGAACATTCTCCCTTTGACGAC 248

RESULT 4

US-09-796-692-5490

Sequence 5490, Application US/09796692

Publication No. US20020198362A1

GENERAL INFORMATION:

APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES

FILE REFERENCE: 2077.001200

CURRENT APPLICATION NUMBER: US/09/796,692

PRIOR FILING DATE: 2001-03-01

PRIOR APPLICATION NUMBER: 60/186,126

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 60/190,479

PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: 60/200,545

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: 60/200,303

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,779

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,999

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: 60/202,084

PRIOR FILING DATE: 2000-05-04

PRIOR APPLICATION NUMBER: 60/206,201

PRIOR FILING DATE: 2000-05-22

PRIOR APPLICATION NUMBER: 60/218,950

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: 60/222,903

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: 60/223,416

PRIOR FILING DATE: 2000-08-04

PRIOR APPLICATION NUMBER: 60/223,378

PRIOR FILING DATE: 2000-08-07

NUMBER OF SEQ ID NOS: 9597

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5490

LENGTH: 326

TYPE: DNA

ORGANISM: Homo sapiens

US-09-796-692-5490

Alignment Scores:

Pred. No.:	0.000424	Length:	326
Score:	86.00	Matches:	19
Percent Similarity:	65.00%	Conservative:	7
Best Local Similarity:	47.50%	Mismatches:	10
Query Match:	23.31%	Indels:	4
DB:	9	Gaps:	2

US-09-935-703-7 (1-67) x US-09-796-692-5490 (1-326)

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

DB 132 GAGGAGTTTGAGAGTTCAGAGGAGGTGAGAACCTTGACACGCGCTG---GAA 188

QY 19 GlyAsnGlnProSerAsnArgLysAsnArgTyrArgAspIleLeuProPheGlnHis 38

DB 189 GGGCAGCGCCAGAGAACAGGCGACAGACCGCTACAGAACATTCTCCCTTTGACGAC 248

RESULT 5

US-10-040-862-3635

Sequence 3635, Application US/10040862

Publication No. US20030078396A1

GENERAL INFORMATION:

APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER

APPLICANT: Retter, Marc

APPLICANT: Corixa Corporation

TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther

TITLE OF INVENTION: Hematological Malignancies

FILE REFERENCE: 014058-013520US

CURRENT APPLICATION NUMBER: US/10/040,862

CURRENT FILING DATE: 2001-11-06

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: US 60/186,126

PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: US 60/190,479

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: US 60/200,545

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: US 60/200,779

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: US 60/200,999

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: US 60/202,084

PRIOR FILING DATE: 2000-05-04

PRIOR APPLICATION NUMBER: US 60/206,201

PRIOR FILING DATE: 2000-05-22

PRIOR APPLICATION NUMBER: US 60/218,950

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: US 60/222,903

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: US 60/223,416

PRIOR FILING DATE: 2000-08-04

PRIOR APPLICATION NUMBER: US 60/223,378

PRIOR FILING DATE: 2000-08-07

PRIOR APPLICATION NUMBER: US 09/796,692

PRIOR FILING DATE: 2001-03-01

NUMBER OF SEQ ID NOS: 10467

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3635

LENGTH: 326

TYPE: DNA

ORGANISM: Homo sapiens

US-10-040-862-3635

Alignment Scores:

Pred. No.:	0.000424	Length:	326
Score:	86.00	Matches:	19
Percent Similarity:	65.00%	Conservative:	7
Best Local Similarity:	47.50%	Mismatches:	10
Query Match:	23.31%	Indels:	4
DB:	9	Gaps:	2

US-09-935-703-7 (1-67) x US-10-040-862-3635 (1-326)

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

DB 132 GAGGAGTTTGAGAGTTCAGAGGAGGTGAGAACCTTGACACGCGCTG---GAA 188

QY 19 GlyAsnGlnProSerAsnArgLysAsnArgTyrArgAspIleLeuProPheGlnHis 38

DB 189 GGGCAGCGCCAGAGAACAGGCGACAGACCGCTACAGAACATTCTCCCTTTGACGAC 248

RESULT 6

US-10-040-862-5490

Sequence 5490, Application US/10040862

Publication No. US20030078396A1

GENERAL INFORMATION:

APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

APPLICANT: Retter, Marc

TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther

TITLE OF INVENTION: Hematological Malignancies

FILE REFERENCE: 014058-013520US

CURRENT APPLICATION NUMBER: US/10/040,862

```

; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5490
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-5490

```

```

Alignment Scores:
Pred. No.: 0.000424      Length: 326
Score: 86.00             Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31%      Indels: 4
DB: 9                     Gaps: 2

```

US-09-935-703-7 (1-67) x US-10-040-862-5490 (1-326)

```

QY 2 GIngluPhemeTaleu-----GluLeuLysAsnLeuProgluYgluPheAsnSer 18
Db 132 GAGGAGTTGAGAGTTTGCAGAGCAGAGGAGTGAAGACTTGCACCGCCTTG--GAA 188
QY 19 GtYanGlnProSerAsnArGtGluYsAsnArGtYrArGAspIleLeuProPheGlnHs 38
Db 189 GGGCAGCGCGCCAGACAGACAGAGGCGCAAGACCGCTACAGAACATCTCCCTTTGACAC 248

```

```

RESULT 7
; Sequence 3721, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR APPLICATION NUMBER: 2001-03-01
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,303

```

```

; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3721
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: n-A,T,C or G
US-09-796-692-3721

```

```

Alignment Scores:
Pred. No.: 0.000436      Length: 333
Score: 86.00             Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31%      Indels: 4
DB: 9                     Gaps: 2

```

US-09-935-703-7 (1-67) x US-09-796-692-3721 (1-333)

```

QY 2 GIngluPhemeTaleu-----GluLeuLysAsnLeuProgluYgluPheAsnSer 18
Db 139 GAGGAGTTGAGAGTTTGCAGAGCAGAGGAGTGAAGACTTGCACCGCCTTG--GAA 195
QY 19 GtYanGlnProSerAsnArGtGluYsAsnArGtYrArGAspIleLeuProPheGlnHs 38
Db 196 GGGCAGCGCGCCAGACAGACAGAGGCGCAAGACCGCTACAGAACATCTCCCTTTGACAC 255

```

```

RESULT 8
; Sequence 3721, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
; FILE REFERENCE: 014058-01352005
; CURRENT APPLICATION NUMBER: US/10/040,862
; PRIOR APPLICATION NUMBER: 2001-11-06
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR APPLICATION NUMBER: US 60/200,999

```

```

; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3721
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: n-A,T,C or G
US-10-040-862-3721

Alignment Scores:
Pred. No.: 0.000436 Length: 333
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
DB: 9 Gaps: 2

US-09-935-703-7 (1-67) x US-10-040-862-3721 (1-333)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsaNleuProGlyGluPheAsnSer 18
Db 139 GAGGAGTTGAGAGTTTCGACAGACGAGGAGTGAGAACTTGACACCGCTCTG---GAA 195

QY 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 196 GGGCAGCGCCGACAGAACAGGCAAGGCAAGCCCTACAGAACTTCTCCCTTTGACAC 255

RESULT 9
US-09-822-849A-5
; Sequence 5, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-5
```

```

Alignment Scores:
Pred. No.: 0.00494 Length: 2076
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
DB: 10 Gaps: 2

US-09-935-703-7 (1-67) x US-09-822-849A-5 (1-2076)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsaNleuProGlyGluPheAsnSer 18
Db 808 GAGGAGTTGAGAGTTTCGACAGACGAGGAGTGAGAACTTGACACCGCTCTG---GAA 864

QY 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 865 GGGCAGCGCCGACAGAACAGGCAAGGCAAGCCCTACAGAACTTCTCCCTTTGACAC 924

RESULT 10
US-09-920-021A-2
; Sequence 2, Application US/09920021A
; Patent No. US20020110800A1
; GENERAL INFORMATION:
; APPLICANT: KAPLAN, DAVID
; APPLICANT: MARSH, H. NICK
; TITLE OF INVENTION: USE OF SHP-1 AND SHP-2 TO DETECT
; FILE REFERENCE: 08338/016001
; CURRENT APPLICATION NUMBER: US/09/920,021A
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: US/08/918,157
; PRIOR FILING DATE: 1997-08-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2145
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-021A-2

Alignment Scores:
Pred. No.: 0.00516 Length: 2145
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
DB: 10 Gaps: 2

US-09-935-703-7 (1-67) x US-09-920-021A-2 (1-2145)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsaNleuProGlyGluPheAsnSer 18
Db 880 GAGGAGTTGAGAGTTTCGACAGACGAGGAGTGAGAACTTGACACCGCTCTG---GAA 936

QY 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 937 GGGCAGCGCCGACAGAACAGGCAAGGCAAGCCCTACAGAACTTCTCCCTTTGACAC 996

RESULT 11
US-09-917-800A-1478
; Sequence 1478, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
```



```

; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1478
; LENGTH: 2176
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 U77038
US-09-917-800A-1478

Alignment Scores:
Pred. No.: 0.00526      Length: 2176
Score: 86.00           Matches: 19
Percent Similarity: 62.50%      Conservative: 6
Best Local Similarity: 47.50%    Mismatches: 11
Query Match: 23.31%           Indels: 4
                                Gaps: 2
                                DB:

US-09-935-703-7 (1-67) x US-09-917-800A-1478 (1-2176)

OY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db 921 GAGGAGTTTGAGACTTTCGCAAAAGCAAGAGGCAAGAACTTGCACCGCTCTG---GAA 977
OY 19 GlysngInProSeAsnArGluLysAsnArGlyTyrArGAspIleLeuProPheGlnHis 38
Db 978 GGGCAGCGGCCGCGAGAACAGACAGAGCAAGACCGCTACAAAGAACATCTCTCCCTTGACCAC 1037

RESULT 12
US-10-072-036-116
; Sequence 116, Application US/10072036
; Publication No. US20030082564A1
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP
; APPLICANT: Sara BJORN
; APPLICANT: Soren TULLIN
; APPLICANT: Kasper ALMHOLT
; APPLICANT: Kurt SCUDDER
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An I
; FILE REFERENCE: 3759-0120P
; CURRENT APPLICATION NUMBER: US/10/072,036
; PRIOR FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/417,197
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 116
; LENGTH: 2532
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-PTP fusion
; NAME/KEY: CDS

```

```

; LOCATION: (1)..(2529)
US-10-072-036-116

Alignment Scores:
Pred. No.: 0.00643      Length: 2532
Score: 86.00           Matches: 19
Percent Similarity: 65.00%      Conservative: 7
Best Local Similarity: 47.50%    Mismatches: 10
Query Match: 23.31%           Indels: 4
                                Gaps: 2
                                DB:

US-09-935-703-7 (1-67) x US-10-072-036-116 (1-2532)

OY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db 742 GAGGAGTTTGAGACTTTCGCAAGCAGAGGTGAAGAACTTGCACCGCTCTG---GAA 798
OY 19 GlysngInProSeAsnArGluLysAsnArGlyTyrArGAspIleLeuProPheGlnHis 38
Db 799 GGGCAGCGGCCGCGAGAACAGACAGAGGCAAGACCGCTACAAAGAACATCTCTCCCTTGACCAC 858

RESULT 14
US-10-002-600-96
; Sequence 96, Application US/10002600
; Patent No. US20020137077A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawks, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 96
; LENGTH: 3220
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Template ID: 293296.19
US-10-002-600-96

```

```

Alignment Scores:
Pred. No.: 0.00884 Length: 3220
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
DB: 12 Gaps: 2

```

US-09-935-703-7 (1-67) x US-10-002-600-96 (1-3220)

```

Oy 2 GluLeuLysAsnLeuProGlyGluPheAsnSer 18
   ::::::::::::::::::::-GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db 924 GAGGAGTTGAGAGTTGCGAAGCAGGAGTGAGAGACTGCACCGCTCTG---GAA 980

```

```

Oy 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleuProPheGlnHis 38
   ::::::::::::::::::::-GlyAsnGlnProSerAsnArgTyrArgAspIleuProPheGlnHis 38
Db 981 GGGCAGCGCGCCAGAGAACAGGCAAGAACCCCTACAGAACATTCCTCCCTTTGACCCAC 1040

```

RESULT 15

```

US-09-918-995-31823
; Sequence 31823, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31823
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(535)
; OTHER INFORMATION: n - A,T,C or G
US-09-918-995-31823

```

Alignment Scores:

```

Pred. No.: 0.00392 Length: 535
Score: 81.50 Matches: 16
Percent Similarity: 67.74% Conservative: 5
Best Local Similarity: 51.61% Mismatches: 9
Query Match: 22.09% Indels: 1
DB: 9 Gaps: 1

```

US-09-935-703-7 (1-67) x US-09-918-995-31823 (1-535)

```

Oy 8 GluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGlnProSerAsnArgGluLys 2'
   ::::::::::::::::::::-GluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGlnProSerAsnArgGluLys 2'
Db 51 GAGGTGAGAGAACATTGCGACCGCTCTG---GAGGGCGAGCGGCCGAGAGAACAGGGCAG 107

```

```

Oy 28 AsnArgTyrArgAspIleuProPheGlnHis 38
   ::::::::::::::::::::-AsnArgTyrArgAspIleuProPheGlnHis 38
Db 108 AACCGCTACAGAACATTCTCCCTTTGACCCAC 140

```

Search completed: June 11, 2003, 14:56:14  
Job time : 120 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 11, 2003, 11:46:42 ; Search time 1170 Seconds

(without alignments)  
1666.571 Million cell updates/sec

Title: US-09-935-703-7

Sequence: 1 MGFMALEKMLPGEFNSGN.....TTFMHSNGAVSLIRYCA 67

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlp  
-Q=/cgm2\_1/USPTO.spool/US09935703/runat\_06062003\_095211\_6154/app-query.fasta.1.263  
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-USERMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-NO\_MMAR -LARGOQUERY -NEG.SCORES=0 -UNAT\_06062003\_095211\_6154 -NCPU=6 -ICPU=3  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_ov:\*  
8: gb\_ov:\*  
9: gb\_ov:\*  
10: gb\_ov:\*  
11: gb\_ov:\*  
12: gb\_ov:\*  
13: gb\_ov:\*  
14: gb\_ov:\*  
15: gb\_ov:\*  
16: gb\_ov:\*  
17: gb\_ov:\*  
18: gb\_ov:\*  
19: gb\_ov:\*  
20: gb\_ov:\*  
21: gb\_ov:\*  
22: gb\_ov:\*  
23: gb\_ov:\*  
24: gb\_ov:\*  
25: gb\_ov:\*  
26: gb\_ov:\*  
27: gb\_ov:\*  
28: gb\_ov:\*

29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	369	100.0	2499	9	BC036539
2	239.5	64.9	1914	6	AX354535
3	231.5	62.7	2786	9	HSMB00374
4	159	43.1	103737	9	AL603966
5	159	43.1	171930	9	AL450334
6	159	43.1	197389	9	AL358791
7	138	37.4	65229	9	AL672108
8	123	33.3	3090	10	D64141
9	88.5	24.0	892	10	RATPTEB
10	88.5	24.0	2155	10	RATPTEB
11	88.5	24.0	2160	9	HSPTPE
12	88.5	24.0	2209	9	HSB315969
13	88.5	24.0	2210	10	MMU40280
14	88.5	24.0	2292	10	MMU62387
15	88.5	24.0	2827	10	MUSPTPE
16	88.5	24.0	3031	9	HSB430580
17	88.5	24.0	4979	10	MMU36758
18	88.5	24.0	5425	6	AX305420
19	88.5	24.0	5425	10	MMU35368
20	88	23.8	1747	6	I23556
21	88	23.8	2122	10	MUSPTPEB
22	88	23.8	2163	10	MUSPTPEB
23	88	23.8	2168	10	BC012660
24	87.5	23.7	173224	2	AC130106
25	86	23.3	2090	9	HGMHCP
26	86	23.3	2121	9	AF178946
27	86	23.3	2143	6	I23559
28	86	23.3	2145	6	I23558
29	86	23.3	2145	6	I23558
30	86	23.3	2145	6	I23558
31	86	23.3	2176	6	AX401802
32	86	23.3	2190	10	RNU77038
33	86	23.3	2229	9	BC007667
34	86	23.3	2229	9	BC002523
35	86	23.3	2227	9	HSPTPE
36	86	23.3	2532	6	A84501
37	86	23.3	2532	6	A84503
38	86	23.3	2562	6	AX427400
39	85	23.0	2741	3	AF015882
40	83.5	22.6	1840	10	MMSHP1503
41	83	22.5	2416	5	XIUI287
42	81.5	22.1	1099	9	HSPTPEG04
43	81.5	22.1	157115	9	AC006512
44	81.5	22.1	222930	9	HSU47924
45	81	22.0	25643	3	US33332

RESULT 1

# ALIGNMENTS

BC036539 Homo sapi  
AX354535 Sequence  
AL050040 Homo sapi  
AL603966 Human DNA  
AL450334 Human DNA  
AL358791 Human DNA  
AL672108 Human DNA  
D64141 Mus musculus  
D78613 Rat mRNA  
D78613 Rat mRNA  
X54134 Human HPT  
AJ315969 Homo sapi  
U40280 Mus musculus  
U62387 Mus musculus  
D33488 Mouse mRNA  
AJ30580 Homo sapi  
U36758 Mus musculus  
AX305420 Sequence  
U35368 Mus musculus  
I23556 Sequence 1  
M80389 Mouse prote  
M88902 Mus musculus  
BC012660 Mus muscu  
AC130106 Rattus no  
M74903 Human hemat  
AF178946 Homo sapi  
I23559 Sequence 7  
I23558 Sequence 5  
M77273 Human tyros  
AX401802 Sequence  
U77038 Rattus norv  
BC007667 Homo sapi  
BC002523 Homo sapi  
X62055 H. sapiens p  
A84501 Sequence 11  
AX427400 Sequence 11  
A84503 Sequence 11  
AF015882 Caenorhab  
AF015882 Caenorhab  
U65953 Mus musculus  
U15287 Xenopus lae  
U15534 Human prote  
AC006512 Homo sapi  
U47924 Human chrom  
US33332 Caenorhabd1

BC036539  
 LOCUS BC036539 2499 bp mRNA linear PRI 20-AUG-2002  
 DEFINITION Homo sapiens, similar to protein-tyrosine-phosphatase homolog DKFZP566K0524.1 - human (fragment), clone MGC:33863 IMAGE:5262458, mRNA, complete cds.  
 ACCESSION BC036539  
 VERSION BC036539.1 GI:22328116  
 KEYWORDS MGC.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 2499)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 REMARK NHI-MGC Project URL: http://mgc.ncl.nih.gov  
 COMMENT Contact: MGC help desk  
 Email: egapds-remail.nih.gov  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Institute for Systems Biology  
 http://www.systemsbio.org  
 Contact: amandahsystemsbiology.org  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov  
 Series: IRK Plate: 47 Row: 9 Column: 17  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genescan gene prediction, Similarity but not identity to protein.

FEATURES  
 source  
 Location/Qualifiers  
 1..2499  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MGC:33863 IMAGE:5262458"  
 /tissue\_type="brain, hippocampus"  
 /clone\_lib="NHI\_MGC\_95"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"  
 192..665  
 /codon\_start=1  
 /product="similar to protein-tyrosine-phosphatase homolog DKFZP566K0524.1 - human (fragment)"  
 /protein\_id="AAH36539.1"  
 /db\_xref="GI:22328117"  
 /translation="MSSPRDRAEPVNDYEGNDSEADLNREPETLPSSQENTPNSKV FENYVNSEKVKLSLRNPFNDYEDYEEPSGSGSPSMKTATGPRRDWSSDEDEAA GPSQALSPILSGTKIVSEGLDQAQIRPLIFNHEQTAKDKCKLIEKNSV"

BASE COUNT  
 ORIGIN  
 766 a 438 c 486 g 809 t

Alignment Scores:  
 Pred. No.: 7.77e-36 Length: 2499  
 Score: 369.00 Matches: 67  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0

US-09-935-703-7 (1-67) x BC036539 (1-2499)

QY 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 670 ATGCAGCAATTATAGCTTAGAATCTGACCTGCTGAGCTCACTCGGAAAT 729  
 QY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHisGly 40  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 730 CAACCAAGCAACAGAGAAAAAACAGATACCGAGATATCTCCATTTCAACATCATGGA 789  
 QY 41 TyrSerGlyProAsnGluArgThrPheTrpHisGlySerAsnGluGlyAlaValSer 60  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 790 TATAGTGGCCCAATGAGAACACACGCTTGCGATGGTTCAACGAAGAGCATATCA 849  
 QY 61 LeuLeuLeuArgTyrCysAla 67  
 |||||||||||||||||||||||  
 Db 850 CTTTGTTCAGATATGTGCT 870

RESULT 2  
 AX354535 1914 bp DNA linear PAT 06-FEB-2002  
 LOCUS AX354535  
 DEFINITION Sequence 13 from Patent WO0196546.  
 ACCESSION AX354535  
 VERSION AX354535.1 GI:18619338  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1  
 AUTHORS Au-Yang, J., Baughn, M.R., Ding, L., Elliott, V.S., Gandhi, A.R., Griffin, J.A., Hatfield, A., Kearney, L., Lee, E.A., Lu, Y., Nguyen, D.B., Patterson, C., Ramkumar, J., Reddy, R., Sanjivani, M.S., Stewart, E.A., Tang, Y.T., Thornton, M., Tribouley, C.M., Yang, J., Yao, M.G. and Yue, H.  
 TITLE Protein phosphatases  
 JOURNAL Patent: WO 0196546-A 13 20-DEC-2001;  
 Incyte Genomics, Inc. (US)

FEATURES  
 source  
 Location/Qualifiers  
 1..1914  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /note="Incyte ID No: 7476861CB1"

BASE COUNT  
 ORIGIN  
 565 a 378 c 438 g 533 t

Alignment Scores:  
 Pred. No.: 5.78e-20 Length: 1914  
 Score: 239.50 Matches: 66  
 Percent Similarity: 26.29% Conservative: 0  
 Best Local Similarity: 26.29% Mismatches: 1  
 Query Match: 64.91% Indels: 185  
 Gaps: 1

US-09-935-703-7 (1-67) x AX354535 (1-1914)

QY 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 680 ATCCAGGAATTTATGCTTAGAATCTGAGATCTGCTGAGTTCAACTCTGGGAAAT 739  
 QY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuPro 35  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 740 CAACCAAGCAACAGAGAAAAAACAGATACCGAGATATCTTCATATGATTCAACAGCC 799  
 QY 35 35  
 Db 800 GTTCCTCTTGAAAAAGCAAGACATCATGCTATGCTATATAGTAATGCAATGTGT 859  
 QY 35 35  
 Db 860 GGAGAGAGTATTTTATATCGCTACTCAAGACCACTGCTGAGCAACATAGACTTT 919  
 QY 35 35  
 Db 920 TGGCAATGCTGTGGAAAAATATTCAAATGTTATTCATGATTAACAGAGAGATAGAA 979  
 QY 35 35

```

Db      980 GGTGAATTATCAATGCTACCTACTGGCCATTCTCTGAAAGAACCATTTGGAATTG 1039
QY      35 ----- 35
Db      1040 AAACACTTCGCTGTATTCCTGGAGAACTACCAATACCTCAATATTTTCATTCGAAATG 1099
QY      35 ----- 35
Db      1100 TTTCAGTTGTGAGAAATGCACAGGAACTAGTCACTGTGTAACAGTTGCAGTTCACC 1159
QY      35 ----- 35
Db      1160 AAGTGCCAGACCATGCACTCTGCTCAGACAGATAGCTTCAATAATATTCGTTAT 1219
QY      35 ----- 35
Db      1220 GCAAGAGAGACACCTTACAGGACCATGGTTGTCATGCAAGTGGCCGATAGCCGG 1279
QY      36 -----Phe 36
Db      1280 ACAGGGGTCTTCTATGTGTGATGTCGTCTGTGCAATCGTAAGAACTGTTCAATC 1339
QY      37 GlnHisHisGlyTyrSerGlyProAsnGluArgThrPheThrPheHisGlySerAsnGlu 56
Db      1340 AA-CATCATGATGATATGATGCGCCCAATAGAGAACAGTTCTGCGCATGGTTCAACGAA 1398
QY      57 GlyAlaValSerLeuLeuLeuArgTyrCysAla 67
Db      1399 GAGCAGATACACTTTTGTATACGATATGTGCT 1431

RESULT 3
HSM800374      2786 bp      mRNA      linear      PRI 18-FEB-2000
LOCUS      HSM800374
DEFINITION      Homo sapiens mRNA; cDNA DKFZp566K0524 (from clone DKFZp566K0524);
ACCESSION      AL050040
VERSION      AL050040.1 GI:4884281
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 2786)
AUTHORS      Ansoorge, W., Winkler, U., Mewes, H.W., Gaassenhuber, J. and Wiemann, S.
TITLE      Direct Submission
JOURNAL      Submitted (15-MAY-1999) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
COMMENT      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp566K0524) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.
            location/Qualifiers
FEATURES
    source
        1..2786
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKFZp566K0524"
            /tissue_type="kidney"
            /clone_id="566 (synonym: hfkdz). Vector pAMP1; host
            X1-2blue; sites NotI + SalI"
            /dev_stage="fetal"
            1..1198
            /gene="DKFZp566K0524"
            <1..1198
            /gene="DKFZp566K0524"
            /note="strong similarity to protein-tyrosine-phosphatases"
            /codon_start=2

```

```

/product="hypothetical protein"
/protein_id="CAB43248.1"
/db_xref="GI:4884282"
/db_xref="SPTREMBL:Q9Y406"
/translating="GPNMGCWCSCGTGAGABOQAVFENKYNSEKVKSLRNPHNDY
EDVPEPESGSDPAMTARGPFRDRNRSSEDEEAGSQALESPLISDRKIVSEEL
DIOAIIRPLFNFEHOTAIKDCLTLEETAYDINOEPMALEKRLPEEYSGNPS
NREKRYBDIIPYDSTRVPLGSKDYINASYRIYNGGEYFIATOGPLSTIDDFM
OMVLENSNTAMITREMEGGIIRKCHYPISTLKKPLEKHFVPLENTOILOYFIR
MEOVERKSTGSHSVKIQFTWPHDGHVASDFTKIRIRAKSHLITGPVYHCSAG
IGRTGVFLCDVDFCAIVKDCSFNIDIVAMQREDRSGNVOTREOYHFCYDYLEVL
KLTLD"
polya_signal      2761..2766
polya_site      2776
BASE COUNT      841 a 494 c 550 g 901 t
ORIGIN
Alignment Scores:
Pred. No.:      8,36e-19      Length:      2786
Score:      231.50      Matches:      65
Percent Similarity:      25.908      Mismatches:      0
Best Local Similarity:      25.908      Indels:      185
Query Match:      62.74%      Gaps:      1
DB:
US-09-935-703-7 (1-67) x HSM800374 (1-2786)
QY      1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
Db      413 ATGAGAGATTTATGCTTTAGAACTTAAGAACTCCTGCTGATGTTCTACTGGGAAT 472
QY      21 GlnProSerAsnArgGluLysAsnArgTyrArgPheLeuPro----- 35
Db      473 CAACCAAGCAGACGAGAAAGAAACAGATACCGAGATTTCTCATATGATTCACACGCG 532
QY      35 ----- 35
Db      533 GTTCCCTTGGAAGAAAGCAAGCACTACATCATGCTAGTATATTAGATAGCAATTGT 592
QY      35 ----- 35
Db      593 GAGAGAGACTATTTTATATCGCTACTCAAGACCACTGCTGAGCACCATAGTACTTT 652
QY      35 ----- 35
Db      653 TGGCAATGGTGTGAAATAATTAATCAATGTTATTCATGATTAACAGAGATGGAA 712
QY      35 ----- 35
Db      713 GGTGAATTATCAATAGCTACATTAATGCGCCATTCTCTGAAGAGCATTTGGAATTG 772
QY      35 ----- 35
Db      773 AAACACTTCGCTGTATTCCTGGAGAACTACCAATACCTCAATATTTTCATTCGAAATG 832
QY      35 ----- 35
Db      833 TTTCAGTTGTGAGAAATGCCACGGGAATGACCTCTGTAAACAGTTGCAGTTCACC 892
QY      35 ----- 35
Db      893 AAGTGCCAGACCATGCACTCTGCTCAGACATAGTTCATTAATAATATTCGTTAT 952
QY      35 ----- 35
Db      953 GCAAGAGAGACCATCTTACAGACCATGTTGTTCACTGACAGTGGCCGCAATAGCCGG 1012
QY      36 -----Phe 36
Db      1013 ACAGGGGTCTTCTATGTGTGATGTCGTCTTGTGCCATCGTAAGAGACTGTTCAATC 1072
QY      37 GlnHisHisGlyTyrSerGlyProAsnGluArgThrPheThrPheHisGlySerAsnGlu 56
            |||||

```

Db 1073 AA-CATCATGATATAGTGGCCCAATGAGAGAACACGCTTCTGSCATGTTCAACGAA 1131  
 QY 57 GYAlaValSerLeuLeuLeuArgTyrCysAla 67  
 Db 1132 GGAGCAGATATCACTTTTGTACGATATTTGCT 1164

RESULT 4  
 AL603966/c  
 LOCUS 103737 bp DNA linear PRI 30-JAN-2002  
 DEFINITION Human DNA sequence from clone RP11-164N7 on chromosome 10, complete  
 sequence.  
 ACCESSION AL603966  
 VERSION AL603966.9 GI:18476754  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 103737)  
 AUTHORS Bird,C.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk  
 COMMENT During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-164N7 is from the library RPCI-11.1 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6  
 IMPORTANT: This sequence is not the entire insert of clone  
 RP11-164N7. It may be shorter because we sequence overlapping  
 sections only once, except for a short overlap.  
 The true left end of clone RP11-164N7 is at 1 in this sequence. The  
 true left end of clone RP11-135E1 is at 101738 in this sequence. The  
 location/Qualifiers  
 1. 103737  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-164N7"  
 /clone\_1lb="RPCI-11.1"  
 BASE COUNT 31673 a 22545 c 21304 g 28215 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.43e-08 Length: 103737  
 Score: 159.00 Matches: 30  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.09% Indels: 0  
 DB: 9 Gaps: 0

US-09-935-703-7 (1-67) x AL603966 (1-103737)  
 QY 6 AAlaGluLeuLeuAsnLeuProGlyGluPheAsnSerGlyAsnGlnProSerAsnArg 25  
 Db 96401 GCTTGAAGAACTTAAGAAATCTGCTGTGAGTTCACACTGTGGAAATCAACCAACGACAGCA 96342  
 QY 26 GAluysAsnArgTyrArgAspIleuPro 35  
 Db 96341 GAAAAAAGACAGATACCGAGATATTTCTTCCA 96312

RESULT 5  
 AL450334/c  
 LOCUS 171930 bp DNA linear PRI 21-JUN-2002  
 DEFINITION Human DNA sequence from clone RP11-556L1 on chromosome 10, complete  
 sequence.  
 ACCESSION AL450334  
 VERSION AL450334.15 GI:21540028  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 171930)  
 AUTHORS Heath,P.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk  
 COMMENT During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-556L1 is from the library RPCI-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk

FEATURES  
 source  
 1. 171930  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-556L1"  
 /clone\_1lb="RPCI-11.2"  
 BASE COUNT 54090 a 36289 c 33202 g 48349 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.43e-08 Length: 103737  
 Score: 159.00 Matches: 30  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.09% Indels: 0  
 DB: 9 Gaps: 0

Pred. No.: 5.85e-08 Length: 171930  
 Score: 159.00 Matches: 30  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.09% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-935-703-7 (1-67) x AL450334 (1-171930)  
 Oy 6 ALaLeuGlutLeuLeuLeuLeuProGlyGluPheAsnSerGlyAsnGlnProSerAsnArg 25  
 Db 2178 GCTTGAAGACTTAAAGATCTGCTGAGTTAACTCTGGATCAACACACACAGCA 2119  
 Oy 26 GtUtyAsnArgTyrArgAspIleLeuPro 35  
 Db 2118 GAAAAAACAAGATACCGAGATATCTTCCA 2089  
 RESULT 6  
 AL358791/c 197389 bp DNA linear PRI 06-JUN-2001  
 LOCUS Human DNA sequence from clone RP11-42B19 on chromosome 10, complete  
 DEFINITION sequence.  
 ACCESSION AL358791  
 VERSION AL358791.24 GI:14141350  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 197389)  
 AUTHORS Clark, S.  
 JOURNAL Direct Submission  
 Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On May 17, 2001 this sequence version replaced GI:13751420.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; Sw:  
 SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-42B19 is from the library RPCR-11.1 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6  
 This sequence is the entire insert of clone RP11-42B19.  
 FEATURES  
 source  
 1. 197389  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-42B19"  
 /clone\_1lb="RPCR-11.1"  
 6. 370  
 /note="match: GSS: Em:AQ116178"  
 misc\_feature  
 26. 180  
 repeat\_region

misc\_feature  
 76. 566  
 /note="match: GSS: Em:AQ116242"  
 /note="match: GSS: Em:AQ726140"  
 1049. 1114  
 /note="6 copies 11 mer 83% conserved"  
 1192. 1246  
 /note="5 copies 11 mer 87% conserved"  
 1246. 1660  
 /note="12 repeat: matches 2256. 2667 of consensus"  
 2964. 3011  
 /note="1133 repeat: matches 575. 621 of consensus"  
 3170. 3580  
 /note="MSTA repeat: matches 1. 426 of consensus"  
 4275. 4345  
 /note="MIR repeat: matches 106. 181 of consensus"  
 5259. 5374  
 /note="MIR repeat: matches 111. 227 of consensus"  
 5414. 5454  
 /note="12 repeat: matches 2657. 2697 of consensus"  
 5633. 6052  
 /note="TTR12 repeat: matches 1. 342 of consensus"  
 6053. 6357  
 /note="AluSq repeat: matches 1. 304 of consensus"  
 6358. 6450  
 /note="TTR12 repeat: matches 342. 440 of consensus"  
 9115. 9246  
 /note="12 copies 11 mer 65% conserved"  
 9122. 9261  
 /note="35 copies 4 mer aaga 66% conserved"  
 11013. 11192  
 /note="MER45 repeat: matches 1. 178 of consensus"  
 11231. 11638  
 /note="match: GSS: Em:AQ210526"  
 11616. 12203  
 /note="MER4B repeat: matches 1. 611 of consensus"  
 12204. 12435  
 /note="12 repeat: matches 2251. 2500 of consensus"  
 12452. 12704  
 /note="MIR repeat: matches 2. 262 of consensus"  
 12796. 12883  
 /note="12 repeat: matches 2641. 2734 of consensus"  
 12939. 13288  
 /note="match: GSS: Em:A2694508"  
 13577. 13719  
 /note="FLAMC repeat: matches 1. 143 of consensus"  
 13747. 13866  
 /note="MER93 repeat: matches 110. 230 of consensus"  
 13895. 14079  
 /note="MER93 repeat: matches 215. 397 of consensus"  
 14090. 14220  
 /note="MER9B repeat: matches 25. 178 of consensus"  
 14657. 14971  
 /note="AluSc repeat: matches 1. 309 of consensus"  
 15219. 16945  
 /note="LMD repeat: matches 1905. 3500 of consensus"  
 16946. 17256  
 /note="AluY repeat: matches 1. 311 of consensus"  
 17257. 18226  
 /note="LMD repeat: matches 977. 1905 of consensus"  
 18902. 19076  
 /note="LMD repeat: matches 15. 174 of consensus"  
 20392. 20519  
 /note="MTR11 repeat: matches 6. 129 of consensus"  
 20677. 20808  
 /note="FLAMC repeat: matches 2. 133 of consensus"  
 20833. 20868  
 /note="MTR11 repeat: matches 308. 346 of consensus"  
 20869. 21214  
 /note="MTR1A1 repeat: matches 1. 365 of consensus"  
 21217. 21621  
 /note="MTR1A1 internal repeat: matches 927. 1333 of  
 consensus"  
 22088. 22592  
 repeat\_region

```

/note="LTR1A1-internal repeat: matches 160. .712 of
consensus"
repeat_region 22617. .22732
/note="LIMB3 repeat: matches 6022. .6145 of consensus"
repeat_region 22795. .22911
/note="LTR1A1-internal repeat: matches 5. .121 of
consensus"
repeat_region 22917. .23273
/note="LTR1A1 repeat: matches 1. .365 of consensus"
repeat_region 23374. .23344
/note="LTR1A1 repeat: matches 339. .410 of consensus"
repeat_region 24336. .24385
/note="LTR1F repeat: matches 268. .317 of consensus"
repeat_region 25059. .26364
/note="LIM4 repeat: matches 1064. .2552 of consensus"
repeat_region 26398. .26843
/note="LIM4 repeat: matches 2651. .3102 of consensus"
repeat_region 26879. .27248
/note="LIMB3 repeat: matches 5551. .5956 of consensus"
repeat_region 27255. .27500
/note="LTR1F repeat: matches 304. .568 of consensus"
repeat_region 27667. .27736
/note="LTR1F repeat: matches 3. .74 of consensus"
repeat_region 27759. .27927
/note="LIMB3 repeat: matches 5992. .6162 of consensus"
repeat_region 30024. .30310
/note="LIME repeat: matches 1151. .1435 of consensus"
repeat_region 31602. .31678
/note="MIR repeat: matches 79. .155 of consensus"
misc_feature 31627. .31927
/note="match: STS: Em:HSJ10C10"
repeat_region 31786. .31900
/note="MIR repeat: matches 89. .203 of consensus"
repeat_region 32440. .32738
/note="AlusX repeat: matches 1. .300 of consensus"
repeat_region 32827. .32870
/note="22 copies 2 mer lt 75% conserved"
repeat_region 34892. .35012
/note="L2 repeat: matches 2584. .2710 of consensus"
repeat_region 36225. .36549
/note="MSRD repeat: matches 21. .394 of consensus"
repeat_region 36732. .36915
/note="MIR repeat: matches 36. .234 of consensus"
repeat_region 37945. .38060
/note="MIR repeat: matches 48. .170 of consensus"
repeat_region 38422. .38524
/note="MERSA repeat: matches 9. .116 of consensus"
repeat_region 38641. .38785
/note="MIR repeat: matches 34. .192 of consensus"
repeat_region 39515. .39846
/note="LIMB3 repeat: matches 5845. .6182 of consensus"
repeat_region 40561. .41030
/note="LTR1H repeat: matches 72. .547 of consensus"
repeat_region 41704. .41789
/note="MIR repeat: matches 84. .166 of consensus"
repeat_region 41923. .42236
/note="LIME1 repeat: matches 5808. .6163 of consensus"
repeat_region 45114. .45484
/note="LTR1A1 repeat: matches 1. .365 of consensus"
repeat_region 46495. .46591
/note="LTR1J repeat: matches 43. .140 of consensus"
repeat_region 47474. .47516
/note="L2 repeat: matches 2658. .2700 of consensus"
repeat_region 48316. .48775
/note="LIMC4 repeat: matches 6443. .6923 of consensus"
repeat_region 49375. .49685
/note="AlusX repeat: matches 1. .299 of consensus"
repeat_region 49686. .49720
/note="LTR12 repeat: matches 445. .479 of consensus"
repeat_region 49691. .49724
/note="PTR5 repeat: matches 173. .206 of consensus"
repeat_region 49725. .50193
/note="LTR12 repeat: matches 19. .1523 of consensus"

```

## Alignment Scores:

```

Pred. No.: 6.76e-08 Length: 197389
Score: 159.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.09% Indels: 0
DB: 9 Gaps: 0

```

US-09-935-703-7 (1-67) x AL358791 (1-197389)

QY 6 AlaleuGlubeuysAsnleupProglYgluPheAsnSerGlyAsnGlnProSerAsnArg 25

Db 178010 GCTTGAACCTTAAGATCTGCTGTAAGTCACTCTGGATCAACCAACGACACGA 177951

QY 26 GtUlysAsnAryTyrArGAspIleuPro 35

Db 177950 GAAAAAATAGATACCGAGATATCTTCCA 177921

## RESULT 7

AL672108/c 65229 bp DNA linear PRI 20-AUG-2002

LOCUS Human DNA sequence from clone RP11-142117 on chromosome 10,

DEFINITION complete sequence.

ACCESSION AL672108

VERSION AL672108.7

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE 1 (bases 1 to 65229)

JOURNAL Clark,S

COMMENT Direct Submission

Submitted (20-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk

On Aug 22, 2002 this sequence version replaced g1:22415876.

Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquerry@sanger.ac.uk](mailto:humquerry@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em: EMBL; SW:

SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP

database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/Wormpep](http://www.sanger.ac.uk/Projects/C_elegans/Wormpep)

This sequence

was generated from part of bacterial clone contigs of human

chromosome 10, constructed by the Sanger Centre Chromosome 10

Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr10>

RP11-142117 is from the library RBC1-11.1 constructed by the group

of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACs3.6.

Location/Qualifiers

1. .65229

/organism="Homo sapiens"

FEATURES

source



/db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-142117"  
 /clone\_lib="RP11-11.1"  
 BASE COUNT 19565 a 14038 c 1382 g 18244 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 8 24e-06 Length: 65229  
 Score: 138.00 Matches: 23  
 Percent Similarity: 83.87% Conservative: 3  
 Best Local Similarity: 74.19% Mismatches: 5  
 Query Match: 37.40% Indels: 0  
 Gaps: 0

US-09-935-703-7 (1-67) x AL672108 (1-65229)

OY 30 TYRARGAPLLEuLeuProPhgNHSHISGLYTYRSerGlyProAsnGluArgThr 49  
 Db 33435 TACTGTCACTTATACAGTCAACATCATGATAGTACGCCCAATGAGACACACG 33376  
 OY 50 PHETPHISGLYSerAsnGluAlaValSer 60  
 Db 33375 TTCTGCATGTTTCAACAGCATGACTTCA 33343

RESULT 8  
 D64141 3090 bp mRNA linear ROD 13-FEB-1999  
 LOCUS D64141  
 DEFINITION Mus musculus mRNA for protein-tyrosine-phosphatase, complete cds.  
 ACCESSION D64141.1 GI:2665457  
 VERSION D64141.1 GI:2665457  
 KEYWORDS protein-tyrosine-phosphatase.  
 SOURCE Mus musculus testis cDNA to mRNA.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS 1 (sites)  
 TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 JOURNAL Ohnogi, M., Kuramochi, S., Matsuda, S., and Yamamoto, T.  
 MEDLINE Molecular cloning and characterization of a novel cytoplasmic  
 REFERENCE protein-tyrosine phosphatase that is specifically expressed in  
 AUTHORS spermatocytes  
 TITLE J Biol. Chem. 272 (52), 33092-33099 (1997)  
 JOURNAL 98070510  
 2 (bases 1 to 3090)  
 Ohnogi, M.  
 REFERENCE Direct Submission  
 AUTHORS Submitted (15-SEP-1995) Minho Ohnogi, The University of Tokyo,  
 JOURNAL Department of Oncology, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108,  
 Japan (E-mail: mohns@leims.u-tokyo.ac.jp, Tel: 03-5449-5305,  
 Fax: 03-5449-5413)

FEATURES  
 source  
 location/Qualifiers  
 1..3090

## CDS

/organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /tissue\_type="testis"  
 218..1498  
 /EC\_number="3.1.3.48"  
 /codon\_start=1  
 /product="protein-tyrosine-phosphatase"  
 /protein\_id="BA023761.1"  
 /translation="MSSPRKVGKTRGNDDEEGNSGNLNLNLSPPSSQKPTPKPI  
 FGNKMSENVKPSHLLSPDYEYVPEPLESDTDEVDVDRSLRNWNSMDSSTA  
 GSKTVPVLSGSRSLSKDTEVSSEKELTOAOTRPLIFNSASAMDCNTLOK  
 EELDIIREPLELQMTLPDPSNGTQNRKRLRPSTYRPLCKKNDYINAS  
 YIRVNHSEEFYATAGSPUPETEDFQNVLENKNTAMITREIECGVICYSYTP  
 ISLKEPLEFEHFVLETFHVTQFTVVFQIVKSKTSQCVKHLQFTKMPDHDGPA  
 SADFFIKYRVKSHITGPLLVHCSAGVGRGVICDVVFSALTEKNSEDFINIVT  
 QMRKRGCMIOIKEDYOFCYEIVLEVLQNLALTY"

BASE COUNT 904 a 603 c 647 g 936 t  
 ORIGIN

Alignment Scores:

Pred. No.: 2.35e-05 Length: 3090  
 Score: 123.00 Matches: 26  
 Percent Similarity: 66.00% Conservative: 7  
 Best Local Similarity: 52.00% Mismatches: 16  
 Query Match: 33.33% Indels: 1  
 Gaps: 0

US-09-935-703-7 (1-67) x D64141 (1-3090)

OY 1 MetGngluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20  
 Db 713 ATCCGTGAGTTTGTGAGTTTGAACAAATGACTGCTGCATGACTTCAATTCGGGAAT 772  
 OY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPhgNHSHISGL 40  
 Db 773 AACTACAGACAGAGATAGAACAGATACGAGATATCTTCATATGATTCACACACT 832  
 OY 40 YTYRSerGlyProAsnGluArgThr 49  
 Db 833 GTTCCTTGGAAAAACAGAGACTACA 860

RESULT 9  
 RATPTPEB 892 bp mRNA linear ROD 05-FEB-1999  
 LOCUS RATPTPEB  
 DEFINITION Rat mRNA for protein tyrosine phosphatase epsilon M, partial cds.  
 ACCESSION D78613  
 VERSION D78613.1 GI:1228943  
 KEYWORDS protein tyrosine phosphatase epsilon M; PTP epsilon M.  
 SOURCE Rattus norvegicus Spleen cDNA to mRNA.  
 ORGANISM Rattus norvegicus

REFERENCE  
 AUTHORS 1 (bases 1 to 892)  
 TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 JOURNAL Nakamura, K., Mizuno, Y., and Kikuchi, K.  
 MEDLINE Molecular cloning of a novel cytoplasmic protein tyrosine  
 REFERENCE phosphatase PTP epsilon  
 AUTHORS Blochem. Biophys. Res. Commun. 218 (3), 726-732 (1996)  
 TITLE 96158952  
 2 (bases 1 to 892)  
 Nakamura, K., Mizuno, Y., and Kikuchi, K.  
 REFERENCE Transmembrane and cytoplasmic protein tyrosine phosphatases PTP  
 AUTHORS epsilon generated from a single gene  
 JOURNAL unpublished  
 3 (bases 1 to 892)  
 Nakamura, K.  
 REFERENCE Direct Submission  
 AUTHORS Submitted (07-DEC-1995) Koji Nakamura, Hokkaido University,  
 JOURNAL Institute of Immunological Science, Kita-15, Nishi-7, Kita-ku,  
 Sapporo, Hokkaido 060, Japan (Tel: 011-707-6839, Fax: 011-707-6839)

FEATURES  
 source  
 location/Qualifiers  
 1..892

## CDS

/organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /tissue\_type="Spleen"  
 1..>892  
 /codon\_start=1  
 /product="protein tyrosine phosphatase epsilon M"  
 /protein\_id="BA011433.1"  
 /db\_xref="GI:1228944"  
 /translation="MEPLCPLLASFSLSIATAGQNDTTPTESNMTSTTACGPPDRC  
 SOLPLTWLLPILLLEFLAAYPRFKORAKAYVNSDKKPKNLTLEQEQRMVLS  
 RSSGGRKRYPIYVEHEEETRNASDCKRFEENSLPSGHIQTTELANKKEENRE  
 KNYVPNTLPDHCRLVILSDGJPCSDYINASYIDYKRNKRFIAAGPKQETVNDPW  
 RMYWEDRSATIVMTNLKERKEKCYOWPDQCGTYGNIIRVCVEDCVLVLDYTIKRF  
 CIPOLPDSCKARLVSQLHF"

## BASE COUNT

sig\_peptide 1..54  
 BASE COUNT 245 a 240 c 228 g 179 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 0.115 Length: 892  
 Score: 88.50 Matches: 24



```

CDS
mRNA
/gene="HPTP epsilon"
1..2160
/gene="HPTP epsilon"
52..2154
/gene="HPTP epsilon"
/EC_number="3.1.3.48"
/codon_start=1
/product="protein-tyrosine phosphatase"
/protein_id="CA38069.1"
/db_xref="GI:35792"
/db_xref="SWISS-PROT:P23469"
/translation="MEPLCLPLVGFSLPLALANGNETTADSNETTTSPPDPCGAS
OPLAMLLPLLLVLLLAAYFRFRKRAVSTSKKPNGLIEEOORVLL
SRSPGPKYFPIPEVHEEIRIRSDDCQOFREPNLPSGIOTFEELANKENR
EKRRNPILPNDSRVITISOLDIGPCSDYINASYDYEKKKFLAOCQKSTVDF
WRMYEOKSATIVMTNKKERKEKCHQTPQOCWTGNINVCEDCVLVDYTIK
FCIOPOLPGCKAPRLVSOLHFTSMPPDGVPTPIGMLKFLKVTNLNVAHPVIVH
CSAGVGRGTFTVIDAMAMMAHEQKVDFEVSRIKRNQPMQOTDMQYTFYQALL
EYLYGDELDVSSLEKHLQTMHGTTHFDKIGLEEFKLTNVRIMKEMRTGNLPA
NMKARVIOIIPYDFNRVILSMKRGQETDYINAFIDYRQKDYFIATOGPLAHVE
DFMRIMEMKSHITVMTLEVOEREDQCYOWMPTGSGSVTHGEITIEIKNDTISEASI
RDLVTLNQPQAROEQVRVROFHFHGMPEIGTIPAEKGMIDLIAVOKQOQOTGNH
PIYVCSAGAGRTGFTALSNILERVAKGLDVFQAVKSLRLQRPVHVTLEQYERC
YKVVDFIDIFSDYANK"

sig_peptide
52..108
/gene="HPTP epsilon"
109..2151
/gene="HPTP epsilon"
/product="protein-tyrosine phosphatase"
/EC_number="3.1.3.48"
190..258
/gene="HPTP epsilon"
/note="transmembrane region"

misc_feature
578 a 584 c 561 g 437 t
ORIGIN
BASE COUNT 578 a 584 c 561 g 437 t
ORIGIN

Alignment Scores:
Pred. No.: 0.293 Length: 2160
Score: 88.50 Matches: 24
Percent Similarity: 37.35% Conservative: 7
Best Local Similarity: 28.92% Mismatches: 19
Query Match: 23.98% Indels: 33
Gaps: 3

US-09-935-703-7 (1-67) x HSEPTPE (1-2160)
OY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21
Db 460 GAGAGGTTCAACCTATGCGCATCTGGACACATACAAAGAACTTTTGAACATGCAATAA 519
OY 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProGlyGlnHis 38
Db 520 GAAGAAACAGAGAAAAAACAAGATATCCCAACATCTCCCAATGACCATTTCTAGGGTG 579
OY 39 -----His 39
Db 580 ATTCTGAGCCAACTGATGATGATTCCTGTTACAGACTACATCAATGCTTCTACATAGAT 639
OY 40 GlyTyr-----SerGlyProAsnGluArgThr 49
Db 640 GGTTCACAAAGAGAAATTAATTCATAGCAGCTCAAGGTCCCAACAGAAACGGTTAAC 699
OY 50 ---PheTyr 51
Db 700 GACTTCTGG 708

RESULT 12
LOCUS HSA315969 2209 bp mRNA linear PRI 01-OCT-2001
DEFINITION Homo sapiens mRNA for tyrosine phosphatase epsilon (PTPSE gene),
cytoplasmic splice variant.
ACCESSION AJ315969
VERSION AJ315969.1 GI:15866729

```

```

KEYWORDS
SOURCE alternative splicing; PTPSE gene; tyrosine phosphatase epsilon.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2209)
AUTHORS Mabakken,T.K., Hauge,H., Flime,E.F., Wiedlocha,A. and Aasheim,H.C.
TITLE Expression of human protein tyrosine phosphatase epsilon in
leukocytes, a potential ERK-pathway regulating phosphatase
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 2209)
AUTHORS Mabakken,T.K.
TITLE Direct Submission
JOURNAL Submitted (21-AGO-2001) Mabakken T.K., Immunology, The Norwegian
Radium Hospital, Montebello, N-0310 Oslo, NORWAY
FEATURES
Source
1..2209
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q26"
/cell_line="u937, TPA stimulated"
10..1938
/gene="PTPSE"
10..1938
/gene="PTPSE"
/note="cytoplasmic splice variant"
/evidence="experimental"
/codon_start=1
/product="tyrosine phosphatase epsilon"
/protein_id="CAC8583.1"
/db_xref="GI:15866730"
/translation="MSNRSSFSRLTFRKQKAVSTSDKKPNGLIEEOORVLL
SRSPGPKYFPIPEVHEEIRIRSDDCQOFREPNLPSGIOTFEELANKENR
EKRRNPILPNDSRVITISOLDIGPCSDYINASYDYEKKKFLAOCQKSTVDF
WRMYEOKSATIVMTNKKERKEKCHQTPQOCWTGNINVCEDCVLVDYTIK
FCIOPOLPGCKAPRLVSOLHFTSMPPDGVPTPIGMLKFLKVTNLNVAHPVIVH
CSAGVGRGTFTVIDAMAMMAHEQKVDFEVSRIKRNQPMQOTDMQYTFYQALL
EYLYGDELDVSSLEKHLQTMHGTTHFDKIGLEEFKLTNVRIMKEMRTGNLPA
NMKARVIOIIPYDFNRVILSMKRGQETDYINAFIDYRQKDYFIATOGPLAHVE
DFMRIMEMKSHITVMTLEVOEREDQCYOWMPTGSGSVTHGEITIEIKNDTISEASI
RDLVTLNQPQAROEQVRVROFHFHGMPEIGTIPAEKGMIDLIAVOKQOQOTGNH
PIYVCSAGAGRTGFTALSNILERVAKGLDVFQAVKSLRLQRPVHVTLEQYERC
YKVVDFIDIFSDYANK"

BASE COUNT 651 a 519 c 534 g 505 t
ORIGIN

Alignment Scores:
Pred. No.: 0.3 Length: 2209
Score: 88.50 Matches: 24
Percent Similarity: 37.35% Conservative: 7
Best Local Similarity: 28.92% Mismatches: 19
Query Match: 23.98% Indels: 33
Gaps: 3

US-09-935-703-7 (1-67) x HSA315969 (1-2209)
OY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21
Db 244 GAGAGGTTCAACCTATGCGCATCTGGACACATACAAAGAACTTTTGAACATGCAATAA 303
OY 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProGlyGlnHis 38
Db 304 GAAGAAACAGAGAAAAAACAAGATATCCCAACATCTCCCAATGACCATTTCTAGGGTG 363
OY 39 -----His 39
Db 364 ATTCTGAGCCAACTGATGATGATTCCTGTTACAGACTACATCAATGCTTCTACATAGAT 423
OY 40 GlyTyr-----SerGlyProAsnGluArgThr 49
Db 424 GGTTCACAAAGAGAAATTAATTCATAGCAGCTCAAGGTCCCAACAGAAACGGTTAAC 483

```

QY 50 ---PheTrp 51  
 DB 484 GACTCTGCG 492

RESULT 13  
 MMU40280 2210 bp mRNA linear ROD 12-JUN-1996  
 LOCUS Mus musculus putative protein tyrosine phosphatase mRNA, complete  
 DEFINITION cds.  
 ACCESSION 040280  
 VERSION 040280.1 GI:1373052  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2210)  
 Schmidt, A., Kucledge, S.J., Endo, N., Opas, E., Tanaka, H.,  
 Wesolowski, G., Liu, C.T., Huang, Z., Ramachandaran, C., Rodan, S.B. and  
 Rodan, G.A.  
 Protein-tyrosine phosphatase activity regulates osteoclast  
 formation and function: inhibition by alendronate  
 Proc. Natl. Acad. Sci. U.S.A. 93 (7), 3068-3073 (1996)  
 8610153  
 REFERENCE 2 (bases 1 to 2210)  
 Aaronson, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-NOV-1995) Jeff Aaronson, Bioinformatics, Merck & Co.,  
 Inc., 126 East Lincoln Avenue, Rahway, NJ 07065, USA  
 Location/Qualifiers  
 1..2210  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_type="in vitro generated osteoclasts"  
 64..2163  
 /note="putative protein tyrosine phosphatase; human  
 tyrosine phosphatase"  
 /product="protein tyrosine phosphatase"  
 /protein\_id="AA02190.1"  
 /db\_xref="GI:1373053"  
 /translation="MEPCPLLASFSLSLARAQGNDDTTPESNMTSTTGAAPPDGA  
 SPPILWLLPLLLFLLAAYFRFRKRAVSSNDKMPNGILEBOQVWLLS  
 RSPGPKRFPPIVPLEHLEIRVSADCKRFEFNSLPSGHIGTFELANKENRE  
 KNYRNILPNDHCKVILSOVDGIPCSYIDGYNKKNKFTLAOGKPOEVNEM  
 RMVWEOBSATIVMLTNLKERKEKCYOVPDGGCMTGNTIRVCEDVYLDTYIRK  
 CIIHPLDSCAPRLVSQLHFTSPDGVPTPIGMKFLKATLTPSHAGTIVHC  
 SAGVGRGTIVIDAMDMHSEKQVDFEVSRIKRNQRPQWQTDVQYTFIYQALL  
 YILGDELVSLEHRLQTLHSTATHDKIGLEEFRLKLNVRIMKEMKRGMLPAN  
 MKRAVIOIIPYDFNRVILSMKRGQETDIYINAFIDGRQDYFMATOGPLATGED  
 FMRVWEMKSHITVMLTEVQEREDCKCYOVPTEGSGVTHGDIETIKSDTLSEASVR  
 DFLVTFKQPLARQEOVRVROFHFHGPVGIPIAEKGMIDILAAVOKOQOQGNHP  
 ITHVCSAGAGRTGTIALSNILERKABGLDVFQAVKSLRLQRPVHVTLEQEFCT  
 KVVQDIDIRSDYANFR"

BASE COUNT 592 a 569 c 587 g 462 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.3 Length: 2210  
 Score: 88.50 Matches: 24  
 Percent Similarity: 37.35% Conservative: 7  
 Best Local Similarity: 28.92% Mismatches: 19  
 Query Match: 23.98% Indels: 33  
 DB: 10 Gaps: 3

US-09-935-703-7 (1-67) x MMU40280 (1-2210)

QY 2 GlnclupheketAlaLeugluleulysanLeuproglyglupheasnserscllyasngin 21  
 DB 469 GAGAGTTCATTGCGCATCTGACACATACAGAACCTTGAACACTACCAATATAA 528  
 QY 22 ProserasnaagLulysasnaigtyrargaspilteleuprophoglnhis----- 38

DB 529 GAAGAAACAGAGAAAAAAGAGATACCCCAACATTCGCCAATGATCATGCAGAGTG 588  
 QY 39 -----His 39

DB 569 ATTGAGCGAAGTGGATGGAATCCCTGCTGACTACATTAATGCTTCATACATGAT 648  
 QY 40 GlyTYR-----SerGlyProAsnGluArgThr---- 49

DB 649 GGCTACAAAGAAAGAACAAATTCATAGCAGCTCAAGGCCCTTAGCAGAGACAGTGAAT 708  
 QY 50 ---PheTrp 51  
 DB 709 GACTCTGCG 717

RESULT 14  
 MMU62387 2392 bp mRNA linear ROD 24-JUL-1996  
 LOCUS Mus musculus protein tyrosine phosphatase-e (pipe) mRNA, complete  
 DEFINITION cds.  
 ACCESSION 062387  
 VERSION 062387.1 GI:1439604  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2392)  
 Hou, E.W. and Li, S.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JUN-1996) Lab. of Molecular Genetics, National  
 Institute of Environmental Health Science, Alexander Dr., P.O. Box  
 12333, Research Triangle Park, NC 27709, USA  
 Location/Qualifiers  
 1..2392  
 /organism="Mus musculus"  
 /strain="C57 Black/6"  
 /db\_xref="taxon:10090"  
 /clone="cd-1"  
 /tissue\_type="brain and lung"  
 1..2392  
 /gene="ptpe"  
 226..2325  
 /gene="ptpe"  
 /note="similar to human protein tyrosine phosphatase  
 epsilon"  
 /product="protein tyrosine phosphatase-e"  
 /protein\_id="AA04553.1"  
 /db\_xref="GI:1439605"  
 /translation="MEPCPLLASFSLSLARAQGNDDTTPESNMTSTTGAAPPDGA  
 SPPILWLLPLLLFLLAAYFRFRKRAVSSNDKMPNGILEBOQVWLLS  
 RSPGPKRFPPIVPLEHLEIRVSADCKRFEFNSLPSGHIGTFELANKENRE  
 KNYRNILPNDHCKVILSOVDGIPCSYIDGYNKKNKFTLAOGKPOEVNEM  
 RMVWEOBSATIVMLTNLKERKEKCYOVPDGGCMTGNTIRVCEDVYLDTYIRK  
 CIIHPLDSCAPRLVSQLHFTSPDGVPTPIGMKFLKATLTPSHAGTIVHC  
 SAGVGRGTIVIDAMDMHSEKQVDFEVSRIKRNQRPQWQTDVQYTFIYQALL  
 YILGDELVSLEHRLQTLHSTATHDKIGLEEFRLKLNVRIMKEMKRGMLPAN  
 MKRAVIOIIPYDFNRVILSMKRGQETDIYINAFIDGRQDYFMATOGPLATGED  
 FMRVWEMKSHITVMLTEVQEREDCKCYOVPTEGSGVTHGDIETIKSDTLSEASVR  
 DFLVTFKQPLARQEOVRVROFHFHGPVGIPIAEKGMIDILAAVOKOQOQGNHP  
 ITHVCSAGAGRTGTIALSNILERKABGLDVFQAVKSLRLQRPVHVTLEQEFCT  
 KVVQDIDIRSDYANFR"

BASE COUNT 629 a 632 c 652 g 479 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.326 Length: 2392  
 Score: 88.50 Matches: 24  
 Percent Similarity: 37.35% Conservative: 7  
 Best Local Similarity: 28.92% Mismatches: 19  
 Query Match: 23.98% Indels: 33  
 DB: 10 Gaps: 3





## source

1. .775  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Athersys RAGE Library"  
 /cell\_line="HT1080"  
 /note="See 'creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT  
 213 a 144 c 164 g 254 t

Query Match 39.5%; Score 723; DB 12; Length 775;  
 Best Local Similarity 99.0%; Pred. No. 2,2e-127;  
 Matches 759; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

129 ACATCATGATATAGTGGCCCAATGAGACACAGCTTGTGGCATGGTTCAACGAAG 188  
 9 ACATCATGATATAGTGGCCCAATGAGACACAGCTTGTGGCATGGTTCAACGAAG 67  
 189 AGCAGATACCTTTGTAGATATGCTGAGAGTCTTGGAACTTGTGACTTTGG 248  
 68 AGCAGATACCTTTGTAGATATGCTGAGAGTCTTGGAACTTGTGACTTTGG 127  
 249 ATTAAAGAACTTGTGCTGCTCTCACTTGAATTAACCAAGTGGTTGCACCTCTCA 308  
 128 ATTAAAGAACTTGTGCTGCTCTCACTTGAATTAACCAAGTGGTTGCACCTCTCA 187  
 309 TAAAGAACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 368  
 188 TAAAGAACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 247  
 369 TTTATCAGTTATTTCTTTCTTAAAGCTCCCTGAAGGCAATATCATTTGGGGGT 428  
 248 TTTATCAGTTATTTCTTTCTTAAAGCTCCCTGAAGGCAATATCATTTGGGGGT 307  
 429 GATCAGTGTATTTATGATCTTGTGCTGAGCAATATCAAAATACCTTCCCATTTTCCA 488  
 308 GATCAGTGTATTTATGATCTTGTGCTGAGCAATATCAAAATACCTTCCCATTTTCCA 367  
 489 GTGAACACATGTTTACATTAAGATTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548  
 368 GTGAACACATGTTTACATTAAGATTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 427  
 549 GCCCAATTAAGATTTTAAATATATTCATTAAGATTTTAAAGATTTGAGCTGCTGCTGCTGCT 608  
 428 GCCCAATTAAGATTTTAAATATATTCATTAAGATTTTAAAGATTTGAGCTGCTGCTGCTGCT 487  
 609 AGCTGAGATTTTCCAGAGCTTTGTAAAGTCTTAATCTGGGAGAACATTAAGGCCAATTAATC 668  
 488 AGCTGAGATTTTCCAGAGCTTTGTAAAGTCTTAATCTGGGAGAACATTAAGGCCAATTAATC 547  
 669 ATGACCTCTTCCAGAGCTTTTAAAGACATGCTATTCATGTTCTTTTACTAGAGCTGCT 728  
 548 ATGACCTCTTCCAGAGCTTTTAAAGACATGCTATTCATGTTCTTTTACTAGAGCTGCT 607  
 729 TACTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 788  
 608 TACTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 667  
 789 TTTGGACACAAATTTGGCTTCCATTTTCCATTTTCAATTTTCACTGCTGCTGCTGCTGCTGCTGCT 847  
 668 TTTGGACACAAATTTGGCTTCCATTTTCCATTTTCAATTTTCACTGCTGCTGCTGCTGCTGCTGCT 727  
 848 CTAGACCAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 894  
 728 GTAGACCAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 773

RESULT 2  
 B0008756/c

## LOCUS

B0008756 779 bp mRNA linear EST 26-MAR-2002

DEFINITION UI-H-E10-2y1-d-17-0-UI.s1 NCI-CGAP\_E10 Homo sapiens CDNA clone

IMAGE:5840224.3', mRNA sequence.

ACCESSION B0008756

VERSION B0008756.1 GI:19733657

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Journal

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgap@rsh.rockefeller.edu

Tissue Procurement: Dr. Jose Mercuende

CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be found

through the I.M.A.G.E. Consortium/BLN at: http://image.llnl.gov

The following repetitive elements were found in this CDNA

sequence: 1-45, >AT richlow\_complexity (matched complement)

Seq primer: M13 FORWARD

POLYA=yes

FEATURES

source

location/Qualifiers

1..779

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:5840224"

/clone\_lib="NCI-CGAP\_E10"

/tissue\_type="Chondrosarcoma"

/dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies)"

/note="Organ: Left Pelvis; Vector: pTZ19-Pac (Pharmacia)

with a modified polylinker; Site 1: EcoR I; Site 2: Not I;

NCI-CGAP\_E10 is a CDNA library containing the following

tissue(s): Chondrosarcoma. The library was constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested

with Not I, and cloned directionally into pTZ19-Pac

vector. The oligonucleotide used to prime the synthesis of

first strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is ACACCTGCAC.

TAG\_Lib-UI-H-E10

TAG\_Tissue=chondrosarcoma

TAG\_Seq=ACACTTGCAC\*

BASE COUNT

304 a 106 c 92 g 272 t 5 others

ORIGIN

Query Match 39.0%; Score 714; DB 14; Length 779;

Best Local Similarity 98.2%; Pred. No. 1.1e-125;

Matches 751; Conservative 0; Mismatches 10; Indels 4; Gaps 3;

1068 TTACCTAATTAATTAAGTTGATATAGCATTTATTAATTCAGATCATTCATGTA 1127

779 TTACCTAATTAATTAAG-TGATATAGCATTTAT-ATATCTGAGCATTCAT-GTA 724

1128 TCTCTCAATGTTGATGATTTTCAAACTAAGATGATGATGATTTTTCACAGTTC 1187

723 TCTCTCAATGTTGATGATTTTCAAACTAAGATGATGATGATTTTTCACAGTTC 664

1188 CATTAATCAATTTATTTCTTACTTCTCAGCTGTGTAACATTTAGAAAGCAT 1247

663 CATTAATCAATTTATTTCTTACTTCTCAGCTGTGTAACATTTAGAAAGCAT 604

1248 TGGGACCAATTTTGGAAACAGATTCATGATGATGATGATGATGATGATGATGATGATGAT 1307

Db 603 TGGGAACCAATTGGAACCAAGATTCATGATGAAATGGAACCTTCATATCT 544  
 QY 1308 GTTTTGAAGAAGATGCGCATTTATACAGTAATTTATTTAGACCTTGCCCTGACA 1367  
 Db 543 GTTTTGAAGAAGATGCGCATTTATACAGTAATTTATTTAGACCTTGCCCTGACA 484  
 QY 1368 ATTAATAGTATATTTTGGACAAAGAGTCTGTCGACAGCTATACCTAATTAAGCTA 1427  
 Db 483 ATTAATAGTATATTTTGGACAAAGAGTCTGTCGACAGCTATACCTAATTAAGCTA 424  
 QY 1428 TAAACATATGATGAGTCTGTCGACAGCTATACCTAATTAAGCTA 1487  
 Db 423 TAAACATATGATGAGTCTGTCGACAGCTATACCTAATTAAGCTA 364  
 QY 1488 TATTTGAAATTTTATATATCATATTAATCTTAATTTCTAAGGCTTATTTGT 1547  
 Db 363 TATTTGAAATTTTATATATCATATTAATCTTAATTTCTAAGGCTTATTTGT 304  
 QY 1548 TCTTTGGCTGAATGAGTATATTTGAATTTGGTGAATTAATTAATTTCTGTTAA 1607  
 Db 303 TCTTTGGCTGAATGAGTATATTTGAATTTGGTGAATTAATTAATTTCTGTTAA 244  
 QY 1608 ATATATATATGCAAAATATATTTGATTTAAATCAATATGATTTCTGTTAA 1667  
 Db 243 ATATATATATGCAAAATATATTTGATTTAAATCAATATGATTTCTGTTAA 184  
 QY 1668 TTTATATGATATATATCTGTTAAATTTCTAATTTGATTTGATTTGAT 1727  
 Db 183 TTTATATGATATATATCTGTTAAATTTCTAATTTGATTTGATTTGAT 124  
 QY 1728 TCTTAATTTATTTCTCTCTGTTAAATTAATAGCTATATGATTAACATTTTC 1787  
 Db 123 TCTTAATTTATTTCTCTCTGTTAAATTAATAGCTATATGATTAACATTTTC 64  
 QY 1788 TTCAGAGATTTCTAGCTATTTAATAATTAATTTACTGTG 1832  
 Db 63 TTCAGAGATTTCTAGCTATTTAATAATTAATTTACTGTG 19

RESULT 3  
 BGI21935 980 bp mRNA linear EST 30-JAN-2001  
 LOCUS 602351372P1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4446208 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BGI21935  
 VERSION BGI21935.1 GI:12615444  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM10224 row: p column: 17  
 High quality sequence stop: 679.  
 Location/Qualifiers  
 1. 980

FEATURES  
 source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4446208"  
 /clone\_lib="NIH\_MGC\_90"  
 /tissue\_type="adenocarcinoma, cell line"

BASE COUNT 267 a 222 c 200 g 291 t  
 ORIGIN  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.7 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC library."

Query Match 33.5%; Score 613.6; DB 12; Length 980;  
 Best Local Similarity 97.6%; Pred. No. 1.2e-106;  
 Matches 697; Conservative 0; Mismatches 9; Indels 8; Gaps 7;

QY 358 TCGTTCTGCTTATACAGTTATTTCTTTCTTAAAGCTCCCTGAAGGCAATATAT 417  
 Db 1 TCGTTCTGCTTATACAGTTATTTCTTTCTTAAAGCTCCCTGAAGGCAATATAT 60  
 QY 418 TCGCTTGGGGTATCAGTGTCTTATGATCTTGTGACGACATATCAAAATTAATTC 477  
 Db 61 TCGCTTGGGGTATCAGTGTCTTATGATCTTGTGACGACATATCAAAATTAATTC 120  
 QY 478 CACATTTTCCAGTGAACAGATGTTACATTAACAGATTCGAGCTTGGCTTGA 537  
 Db 121 CACATTTTCCAGTGAACAGATGTTACATTAACAGATTCGAGCTTGGCTTGA 180  
 QY 538 GGGATTACAGAGCCCAATTAAGATTTAAATATATCATAGATTTTATTTGAAG 597  
 Db 181 GGGATTACAGAGCCCAATTAAGATTTAAATATATCATAGATTTTATTTGAAG 240  
 QY 598 TGGCTGAGAGAGCTGAGATTTCCAGACTTTGTAGTCTTATTTCTGGAGAACATA 657  
 Db 241 TGGCTGAGAGAGCTGAGATTTCCAGACTTTGTAGTCTTATTTCTGGAGAACATA 300  
 QY 658 GGCATATATCATGACCTTCTCCAGCATTTTAAAGAGATCTATTCATGTTCTTA 717  
 Db 301 GGCATATATCATGACCTTCTCCAGCATTTTAAAGAGATCTATTCATGTTCTTA 360  
 QY 718 GCTAGAGCTGTACTTTTGTGCGCATTTGAATTAACAGATTTTAAAGAGTCCAGT 777  
 Db 361 GCTAGAGCTGTACTTTTGTGCGCATTTGAATTAACAGATTTTAAAGAGTCCAGT 419  
 QY 778 GGTGAGCTACTTTGGACACAA-ATTGGCTTCCATTTCTTACATTTTCATGCTGCT 836  
 Db 420 GGTGAGCTACTTTGGACACAAACATTTGGCTTCCATTTCTTACATTTTCATGCTGCT 479  
 QY 837 TCTTACAGCTGCTAGACCAAGACCTTGTGTCGGAGACATTTCAATGATAGGAGAGC 896  
 Db 480 TCTTACAGCTGCTAGACCAAGACCTTGTGTCGGAGACATTTCAATGATAGGAGAGC 539  
 QY 897 TCCCTCGGTGAACAGTCCCAAACTAAT-ATGATGTTTATATAGAAG-CCCAAGAGA 954  
 Db 540 TCCCTCGGTGAACAGTCCCAAACTAATATGATGTTTATATAGAAG-CCCAAGAGA 599  
 QY 955 GACTTTGGCATGCTGAGTCTTCTTCCATTCACACCTTACACATTAATTAATTAAT 1014  
 Db 600 GACTTTGGCATGCTGAGTCTTCTTCCATTCACACCTTACACATTAATTAATTAAT 658  
 QY 1015 CTGCTTTGTAAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1068  
 Db 659 CTGCTT-GTTAAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 709

RESULT 4  
 B0787020 607 bp mRNA linear EST 26-JUL-2002  
 LOCUS 1149a08.y1 HR85 islet Homo sapiens cDNA clone IMAGE:603302 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B0787020  
 VERSION B0787020.1 GI:21995492  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 607)  
Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K.,  
Lemishka, I., Scarce, M., Brestell, J., Gradow, G., Clifton, S.,  
Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Birstein, A.,  
Schmitt, A., Theising, B., Rutter, E., Ronko, I., Bennett, J., Cardenas,  
M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, R., Williams, T.,  
Jackson, Y., and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Other ESTs: 1149408.x1  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biochem.harvard.edu  
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Seq primer: -408P from Gibco  
High quality sequence stop: 485.  
Location/Qualifiers

FEATURES  
source

1. 607  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:603302"  
/clone\_1b="HR81818"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
NotI; Site\_2: XhoI; cDNA made by oligo-dT priming.  
Size-selected on agarose gel. Average insert size ~1kb. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permut Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110. E-mail: hinoue@imgate.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-2692."

BASE COUNT  
ORIGIN

Query Match 31.2%; Score 572; DB 14; Length 607;  
Best Local Similarity 99.8%; Pred. No. 1e-98;  
Matches 583; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1105 AATAATTCAGATCATGATGATCTCTCATGTTGATGATATTTTCAAACTAAGATCTA 1164  
1106 AATAATTCAGATCATGATGATCTCTCATGTTGATGATATTTTCAAACTAAGATCTA 1164  
24 AATAATTCAGATCATGATGATCTCTCATGTTGATGATATTTTCAAACTAAGATCTA 83  
1165 TGATAG-TTTTTTCCAGAGTTCATTAATCATTTATTTCTTACTTGCACCTCT 1223  
84 TGATAGTTTTTTTCCAGAGTTCATTAATCATTTATTTCTTACTTGCACCTCT 143  
1224 GTTGAACATTTAGAACTGATTTGGAAACCAATTTTGGAAACCAAGATCATAGTCA 1283  
1225 GTTGAACATTTAGAACTGATTTGGAAACCAATTTTGGAAACCAAGATCATAGTCA 1283  
144 GTTGAACATTTAGAACTGATTTGGAAACCAATTTTGGAAACCAAGATCATAGTCA 203  
1284 TGAANAATGGAACCTCCATATCTGTTTTTGAAGAAAGTGGCCATTTACAGTAATTT 1343  
204 TGAANAATGGAACCTCCATATCTGTTTTTGAAGAAAGTGGCCATTTACAGTAATTT 263  
1344 TATATAGAGACTTCTCGTCAATTAATAGTATTTTGGCAAGAGTCTGCTGA 1403  
264 TATATAGAGACTTCTCGTCAATTAATAGTATTTTGGCAAGAGTCTGCTGA 323  
1404 CAAGTATACCTAATTAATAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 1463  
324 CAAGTATACCTAATTAATAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 383

QY 1464 CATGAGATCAGATATCTATGATTTGAGAAAGCTTTAATCATCTAATCTT 1523  
DB 384 CATTGAGATCAGATATCTATGATTTGAGAAAGCTTTAATCATCTAATCTT 443  
QY 1524 GAATTTCTAAGAGCTTATTTGTTTGGCTGATGATGATTAATTTGAATGTTGAA 1583  
DB 444 GAATTTCTAAGAGCTTATTTGTTTGGCTGATGATGATTAATTTGAATGTTGAA 503  
QY 1584 TATTTAATTAATTTCTCTTTGTAATAATTAATTAATTAATTAATTTGATGTTAAATC 1643  
DB 504 TATTTAATTAATTTCTCTTTGTAATAATTAATTAATTAATTAATTTGATGTTAAATC 563  
QY 1644 AATGATGATCTCTGTTTCAATGTCATTAATTAATTAATTAATTTGATGTTAAATC 1687  
DB 564 AATGATGATCTCTGTTTCAATGTCATTAATTAATTAATTAATTTGATGTTAAATC 607

RESULT 5  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA443172 575 bp mRNA linear EST 03-JUN-1997  
2X98610.r1 Soares.NbHMPu.S1 Homo sapiens cDNA clone IMAGE:811818  
5', mRNA sequence.  
AA443172 GI:2155847  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 575)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,  
Schellenberg, K., Stepec, M., Tan, F., Theising, B., White, Y., Wylie,  
T., Waterston, R., and Wilson, R.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 453.  
Location/Qualifiers

FEATURES  
source

1. 575  
/organism="Homo sapiens"  
/db\_xref="GDB:6042729"  
/db\_xref="taxon:9606"  
/clone="IMAGE:811818"  
/clone\_1b="Soares.NbHMPu.S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NBH, pregnant uterus  
NbHMPu, and fetal heart NbHMPu) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of T.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

BASE COUNT  
ORIGIN

Query Match 30.3%; Score 554.8; DB 9; Length 575;  
Best Local Similarity 99.6%; Pred. No. 1.9e-95;  
Matches 556; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Seq primer: -400p from Glibco  
High quality sequence stop: 383.  
Location/Qualifiers  
1. 535  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="6033302"  
/clone\_id="HR85 Islet"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site: 1:  
Motif: Site 2: XhoI; CDNA made by oligo-dT priming.  
Size-selected on agarose gel. Average insert size -1kb. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permutt Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-7692."

BASE COUNT 215 a 75 c 56 g 189 t

ORIGIN

Query Match 27.9%; Score 511.8; DB 14; Length 535;  
Best Local Similarity 99.4%; Pred. No. 2.7e-87;  
Matches 524; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1306 CTGTTTGAAGAAAGTGGCCATTTATACAGTAATTTATATAGACTTGGCTCGTA 1365  
DB 535 CTGTTTGAAGAAAGTGGCCATTTATACAGTAATTTATATAGACTTGGCTCGTA 476  
QY 1366 CAATTAATAGCAATTTTGGCAAGAGGCTTGGTGCACAGCTTACCTAATTTAAGC 1425  
DB 475 CAATTAATAGCAATTTTGGCAAGAGGCTTGGTGCACAGCTTACCTAATTTAAGC 416  
QY 1426 TATAAACAATAGATAGTGTGTGTGACAGTTTACCAATGAGATCAGAAATTTCTA 1485  
DB 415 TATAAACAATAGATAGTGTGTGTGACAGTTTACCAATGAGATCAGAAATTTCTA 356  
QY 1486 TGTATGAGAAATGTTTAAATATCAATCTAATTAATCTGAATTTCTAAGAGGCTTATTT 1545  
DB 355 TGTATGAGAAATGTTTAAATATCAATCTAATTAATCTGAATTTCTAAGAGGCTTATTT 296  
QY 1546 GTTCTTTTGGCTGAATGATATTTGAATTTGGTGAATTAATTAATTTCTCTGTA 1605  
DB 295 GTTCTTTTGGCTGAATGATATTTGAATTTGGTGAATTAATTAATTTCTCTGTA 236  
QY 1606 AAATTAATTAATGCAAAAATATATTTGATGTTAAATGAATGATTTCTGTTACAT 1665  
DB 235 AAATTAATTAATGCAAAAATATATTTGATGTTAAATGAATGATTTCTGTTACAT 176  
QY 1666 TGTTCATATGATATATATATCTGTGTTAATTTCAATTTGATATATGAGGCTTAAATTTGT 1725  
DB 175 TGTTCATATGATATATATCTGTGTTAATTTCAATTTGATATATGAGGCTTAAATTTGT 116  
QY 1726 ATCTCTAATTTATTTTCTCTCTGTCTACTGTAATTAATTAATGATATGATTAACATTT 1785  
DB 115 ATCTCTAATTTATTTTCTCTCTGTCTACTGTAATTAATTAATGATATGATTAACAA-TT 57  
QY 1786 TCTTCAGAGAAATCTATGCTATATTAATTAATTAATTAATTTTACCGTG 1832  
DB 56 TCTTCAGAGAAATCTATGCTATATTAATTAATTAATTAATTTTACCGTG 10

RESULT 6  
LOCUS B0786761 535 bp mRNA linear EST 26-JUL-2002  
DEFINITION 1149a08.x1 HR85 Islet Homo sapiens cDNA clone IMAGE:6033302 3',  
mRNA sequence.  
ACCESSION B0786761  
VERSION B0786761.1 GI:21995233  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
REFERENCE 1 (bases 1 to 535)  
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
Lemishka, I., Scaerch, M., Bresnelli, J., Gradwohl, G., Clifford, S.,  
Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Blistein, A.,  
Schmitt, A., Theising, B., Ritzer, E., Ronko, I., Bennett, J., Cardenas,  
M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T.,  
Jackson, Y., and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@oebio.h.harvard.edu  
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center for information on

RESULT 7  
LOCUS AM241948 491 bp mRNA linear EST 14-DEC-1999  
DEFINITION X077604.x1 Soares\_NFL\_T-GBC-St Homo sapiens cDNA clone  
IMAGE:2700486 3', mRNA sequence.  
ACCESSION AM241948  
VERSION AM241948.1 GI:6575702  
KEYWORDS EST.

Source	Organism	Human
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.	
AUTHORS	1 (bases 1 to 491)	
TITLE	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps-r@mail.nih.gov">cgaps-r@mail.nih.gov</a> This clone is available royalty-free through LNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. Seq primer: -40UP from Glbbo High quality sequence stop: 454.	
FEATURES	Location/Qualifiers	
source	1..491	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone IMAGE:2700486"	
	/clone_lib="Soares_NFL_T_GBC_S1"	
	/lab_host="DH10B"	
	/note="Organ: pooled; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NH7, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT	201 a 68 c 49 g 173 t	
ORIGIN		
Query Match	26.4% Score 484.4; DB 10; Length 491;	
Best Local Similarity	99.8%; Pred. NO. 4.3e-82;	
Matches	485; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1347 TATAGGACTTGGCTCGTACAAATTAATGATATTTTGGACAAGAGTTCGGTGACAA	14006
DB	491 TATAGGACTTGGCTCGTACAAATTAATGATATTTTGGACAAGAGTTCGGTGACAA	432
QY	1407 GCTTACCTAATTTAAGCTRTAAACATATGATATGATGTTGTACAGTTTAATCTCA	14666
DB	431 GCTTACCTAATTTAAGCTRTAAACATATGATATGATGTTGTACAGTTTAATCTCA	372
QY	1467 TGGAGATCAGAAATTTCTATGATATGAGAAATTTTAAATTCATATTAATCTTGAA	15286
DB	371 TGGAGATCAGAAATTTCTATGATATGAGAAATTTTAAATTCATATTAATCTTGAA	312
QY	1527 TTTCTAAGAGCTTATTTTGTCTTTTGGCTGATGATATTTGAATTTGGTGAATTA	15866
DB	311 TTTCTAAGAGCTTATTTTGTCTTTTGGCTGATGATATTTGAATTTGGTGAATTA	252
QY	1587 TTAATATTTCTCATGTGAAAAATTAATATATGCCAAAAATATATTTGATGTTAAATCA	16466
DB	251 TTAATATTTCTCATGTGAAAAATTAATATATGCCAAAAATATATTTGATGTTAAATCA	192
QY	1647 TAGATGATTTCTGTTTACATGTTTCATATGATATGATATATATGATATTTTCAATTTGATA	1706
DB	191 TAGATGATTTCTGTTTACATGTTTCATATGATATGATATATATGATATTTTCAATTTGATA	132
QY	1707 ATTGGCCCTTAATATTTGATATCTATATTTTATTTTCTCTGTTACTGTTAAATTAATAG	1766
DB	131 ATTGGCCCTTAATATTTGATATCTATATTTTATTTTCTCTGTTACTGTTAAATTAATAG	72
QY	1767 CTATATATGTTAATCAATTTTCTTCAGAAAGATTCATATGATATTTTAAATTAATATTTT	1826
DB	71 CTATATATGTTAATCAATTTTCTTCAGAAAGATTCATATGATATTTTAAATTAATATTTT	12

QY	1827	ACTGTC	1832
Dd	11	ACTGTC	6
RESULT 8			
BG655441			
LOCUS	b34e04.y1	HR85	islet Homo sapiens cDNA 5', mRNA sequence.
DEFINITION			
ACCESSION	BG655441		
VERSION	BG655441.1	GI:13792850	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match	25.9%	Score 474;	DB 12; Length 499;
Best Local Similarity	99.8%	Pred. No. 4.1e-80;	
Matches 485; Conservative	0;	Mismatches 0;	Indels 1; Gaps 1;
QY	1105	ATAATTTGGATCATTCATGTAGTCTCAGTCTGATTTGAATTTTCAACTAAGATCA	1164
Dd	14	ATAATTGTGATCATTCATGTAATCTCTCAATGTTGATTTTCAAACTAAGATCA	73
QY	1165	TGATAG-TTTTTTTCCAGAGTTCATTAATCATTTATTTCCCTTACTTTCACACTCT	1223
Dd	74	TGATAGTTTTTTTTTCCAGAGTTCATTAATCATTTATTTCCCTTACTTTCACACTCT	133
QY	1224	GTTGAACAATTGGAACACTGGATTTGGGAACCAATTTTGGAACACGATTCATAGTCA	1283
Dd	134	GTTGAACAATTGGAACACTGGATTTGGGAACCAATTTTGGAACACGATTCATAGTCA	193

QY 1284 TGAAGATGAAATTCATATTCGTTTGAAGAAAGATGGCCATTATACAGTAATT 1343  
 Db 194 TGAAGATGAAATTCATATTCGTTTGAAGAAAGATGGCCATTATACAGTAATT 253  
 QY 1344 TATATAGAGACTTGGCTGCTAGCAATATATAGATATTTTGGCAAGAGATTCGTGA 1403  
 Db 254 TATATAGAGACTTGGCTGCTAGCAATATATAGATATTTTGGCAAGAGATTCGTGA 313  
 QY 1404 CAAGCTATACCTATATATAGCTATATAAACAATAGATAGATGTTTGTACAGTTAACT 1463  
 Db 314 CAAGCTATACCTATATATAGCTATATAAACAATAGATAGATGTTTGTACAGTTAACT 373  
 QY 1464 CAATGAGATCAGAAATATCTATATATGAGAAATGTTTAAATCAATCATATAATCT 1523  
 Db 374 CAATGAGATCAGAAATATCTATATATGAGAAATGTTTAAATCAATCATATAATCT 433  
 QY 1524 GAATTTCTAAGAGCTTATTTTGTCTTTGGCTGAATGAGTATATTTGAATGTTGAA 1583  
 Db 434 GAATTTCTAAGAGCTTATTTTGTCTTTGGCTGAATGAGTATATTTGAATGTTGAA 493  
 QY 1584 TAATTA 1589  
 Db 494 TAATTA 499  
 RESULT 9  
 BG655503 396 bp mRNA linear EST 05-JUL-2001  
 LOCUS 1B34E04.X1 HR85 1stet Homo sapiens cDNA 3', mRNA sequence.  
 DEFINITION BG655503  
 VERSION BG655503.1 GI:13792912  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 396)  
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
 Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, D., Blistain, A.,  
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas,  
 M., Gibbons, M., McCann, R., Cole, R., Tsagarashvili, R., Williams, T.,  
 Jackson, Y., and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Other ESTs: 1B34E04.Y1  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8537  
 Email: dmelton@biochem.harvard.edu  
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Hiroshi Inoue  
 (hinoue@im.wustl.edu)  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 337.  
 Location/Qualifiers  
 1. 396  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HR85 1stet"  
 /tissue\_type="Purified pancreatic islet"  
 /lab\_host="DH10B"  
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
 Motif\_Site\_2: XhoI; cDNA made by oligo-dT priming.  
 Size-selected on agarose gel. Average insert size ~1kb. 5'  
 XhoI site was destroyed after directional cloning.  
 Amplified once. Contact information: Hiroshi Inoue, MD,

Metabolism Div. (Alan Permutt Lab), Washington University  
 School of Medicine, Box 8127, 660 South Euclid Ave., St.  
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:  
 314-362-1916, Fax: 314-747-2692.  
 BASE COUNT 161 a 46 c 35 g 154 t  
 ORIGIN  
 Query Match 19.5%; Score 357; DB 12; Length 396;  
 Best Local Similarity 99.7%; Pred. No. 6, 6e-58;  
 Matches 368; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 QY 1463 TCAATGAGATCAGAAATATCTATATATGAGAAATGTTTAAATCAATCATATAATCT 1522  
 Db 396 TCAATGAGATCAGAAATATCTATATATGAGAAATGTTTAAATCAATCATATAATCT 337  
 QY 1523 TGAATTTCTAAGAGCTTATTTTGTCTTTGGCTGAATGAGTATATTTGAATGTTGAA 1582  
 Db 336 TGAATTTCTAAGAGCTTATTTTGTCTTTGGCTGAATGAGTATATTTGAATGTTGAA 277  
 QY 1583 ATAATTAATATTCATGTAATAAATATATATATATATATATATATATATATATATAT 1642  
 Db 276 ATAATTAATATATTCATGTAATAAATATATATATATATATATATATATATATATAT 217  
 QY 1643 CAATATGATGATCTGTTTACATTTGTTCAATGATGATATATATATATATATATATAT 1702  
 Db 216 CAATATGATGATCTGTTTACATTTGTTCAATGATGATATATATATATATATATATAT 157  
 QY 1703 GATTAATGGCTTTTAAATATTTGTTATCTGTTATATTTTCTGTTACTGTAATAATA 1762  
 Db 156 GATTAATGGCTTTTAAATATTTGTTATCTGTTATATTTTCTGTTACTGTTAATAATA 97  
 QY 1763 ATAGCTATATGAT 1822  
 Db 96 ATAGCTATATGAT 38  
 QY 1823 ATTTACTGT 1831  
 Db 37 ATTTACTGT 29  
 RESULT 10  
 AM868331 302 bp mRNA linear EST 22-MAY-2000  
 LOCUS MR1-SN0060-050500-002-f11 SN0060 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION AM868331  
 VERSION AM868331.1 GI:8002383  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 302)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, N. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?rl=st2-MR1-SN0060-050

500-002-fillc3-2000-05-05c14-1)  
Seq primer: puc 18 forward  
High quality sequence start: 61  
High quality sequence stop: 302.  
Location/Qualifiers  
1. .302

## FEATURES

source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="SN0060"  
/dev\_stage="Adult"  
/note="Organ: stomach normal. Vector: puc18; Site.1: Sma1; Site.2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 82 a 68 c 59 g 93 t

## ORIGIN

Query Match 15.6%; Score 286; DB 10; Length 302;  
Best Local Similarity 98.3%; Pred. No. 2e-44;  
Matches 289; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 657 AGGCCAATATCATATACCTCTCCAGGCAATTTTAAGACAGATGCTATTCATGTTCTTT 716  
DB 7 AGGCAACTGTCATACCTCTCCAGGCAATTTTAAGACAGATGCTATTCATGTTCTTT 66  
QY 717 AGCTAGACCTCTGATCTTTTCTGCGATTTGAATACCCAGTTAAAGAGCCAGTTA 776  
DB 67 AGCTAGACCTCTGATCTTTTCTGCGATTTGAATACCCAGTTAAAGAGCCAGTTA 126  
QY 777 GGGTGGACTAATCTTGGACACAAATGGCTTCATTCTCTACATTTTCACTGCTGCT 836  
DB 127 GGGTGGACTAATCTTGGACACAAATGGCTTCATTCTCTACATTTTCACTGCTGCT 186  
QY 837 TCTTACAGCTGCTATACCAAGACCTGTTGGTGGACATTTCTGATAGGAGAGC 896  
DB 187 TCTTACAGCTGCTATACCAAGACCTGTTGGTGGACATTTCTGATAGGAGAGC 246  
QY 897 TCTCTCGGTGAAACGTCCTCAAACTAAATAGATGTTTATATAGAAAGCCCAAG 950  
DB 247 TCTCTCGGTGAAACGTCCTCAAACTAAATAGATGTTTATATAGAAAGCCCAAG 300

## RESULT 11

LOCUS D62174 260 bp mRNA linear EST 29-AUG-1995  
DEFINITION H0M243G02B Clontech human aorta polyA+ mRNA (#6572) Homo sapiens  
ACCESSION D62174  
VERSION D62174.1 GI:965950  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

AUTHORS

1 (bases 1 to 260)  
Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,  
Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaiichi,  
A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H.,  
Shin,S. and Nakamura,Y.  
Fujiwara et al. (1995)  
Unpublished (1995)

## COMMENT

Contact: Tsutomu Fujiwara  
Otsuka Gen Research Institute  
Otsuka Pharmaceutical Co., Ltd  
463-10 Kagasuno Kawachi-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035.

## FEATURES

source

Location/Qualifiers  
1. .260  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone="GEN-243G02"  
/clone\_1lb="Clontech human aorta polyA+ mRNA (#6572)"  
/note="Male adult, hematopoietic tissue, stem cell"

BASE COUNT 87 a 21 c 34 g 109 t

## ORIGIN

Query Match 13.9%; Score 255.2; DB 14; Length 260;  
Best Local Similarity 96.5%; Pred. No. 1.4e-38;  
Matches 251; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

QY 1444 AGTGTGTACAGTTTAACTCAATGAGATGAGATATTCATGATGAGAAATGTTT 1503  
DB 1 AGTGTGTACAGTTTAACTCAATGAGATGAGATATTCATGATGAGAAATGTTT 60  
QY 1504 AATATCAATCTATTAATTTTGAATTTCTAAGAGCTTATTTGTTCTTGGCTAATGA 1563  
DB 61 AATATCAATCTATTAATTTTGAATTTCTAAGAGCTTATTTGTTCTTGGCTAATGA 120  
QY 1564 GATATTTGAATTTGTTGAATTAATTAATTTCTATGTAATAATTAATGCAAA 1623  
DB 121 GATATTTGAATTTGTTGAATTAATTAATTTCTATGTAATAATTAATGCAAA 180  
QY 1624 AATATTTGATGTTTAAATCAATAGATGATTTGTTTACATTTGTCATATGATATTA 1683  
DB 181 AATATTTGATGTTTAAATCAATAGATGATTTGTTTACATTTGTCATATGATATTA 240  
QY 1684 TCTGTGTAATTTCAATTTG 1703  
DB 241 TCTGTGTAATTTCAATTTG 260

## RESULT 12

LOCUS B1460524 571 bp mRNA linear EST 21-AUG-2001  
DEFINITION 603201139F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5267145 5',  
mRNA sequence.

ACCESSION B1460524  
VERSION B1460524.1 GI:15251180  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 571)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: egads-remail.nih.gov

Tissue Procurement: Miklos Palcovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LMN at:

http://image.llnl.gov

Plate: LMN1673 row: n column: 10

High quality sequence stop: 569.

## FEATURES

source

Location/Qualifiers  
1. .571  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5267145"  
/clone\_1lb="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: pBluescript (modified  
pBluescript KS+); Site.1: BamHI; Site.2: SalI; XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.2 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 11:04:12 ; Search time 262 seconds  
(Without alignments)  
9786.078 Million cell updates/sec

Title: US-09-935-703-2

Perfect score: 1832

Sequence: 1 ggcacgaggtatgatacat.....aaataaataattactctg 1832

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 870385 seqs, 699768693 residues 1740770

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	360.6	19.7	381	9	US-10-076-622-208
2	360.6	19.7	381	10	US-09-604-287A-208
3	360.6	19.7	381	10	US-09-339-338-208
4	360.6	19.7	381	12	US-10-007-805-208
5	71.8	3.9	15732	9	US-10-239-676-95
6	70	3.8	5689	9	US-10-239-676-90
7	67.8	3.7	11047	9	US-10-239-676-187
8	67.6	3.7	6030	9	US-10-239-676-164
9	65.4	3.6	7823	9	US-10-239-676-197
10	65.2	3.6	4985	9	US-10-094-240-10
11	64	3.5	502	9	US-10-198-846-8860
12	64	3.5	5979	9	US-10-239-676-18
13	61.6	3.4	6620	9	US-10-239-676-195
14	61.2	3.3	9539	9	US-10-239-676-51
15	61.2	3.3	11047	9	US-10-239-676-188
16	60.6	3.3	3007	9	US-10-239-676-219
17	60.6	3.3	9515	9	US-10-239-676-160
18	60	3.3	15732	9	US-10-239-676-96
19	59.6	3.3	6337	9	US-10-239-676-151

20	59.2	3.2	8588	9	US-10-239-676-177	Sequence 177, App
21	58.8	3.2	169139	9	US-10-067-514-1	Sequence 1, App1
22	58.2	3.2	5979	9	US-10-239-676-18	Sequence 18, App1
23	58	3.2	4985	9	US-10-094-240-10	Sequence 10, App1
24	57.8	3.2	6306	9	US-10-239-676-129	Sequence 19, App1
25	57.4	3.1	11260	9	US-10-239-676-170	Sequence 19, App1
26	57.2	3.1	5610	9	US-10-239-676-170	Sequence 19, App1
27	57.2	3.1	11812	9	US-10-239-676-210	Sequence 210, App
28	57.2	3.1	17421	9	US-10-239-676-53	Sequence 53, App1
29	56.8	3.1	431	10	US-09-960-352-5558	Sequence 5558, App
30	56.2	3.1	516	10	US-09-960-352-5785	Sequence 5785, App
31	56.2	3.1	8866	9	US-10-239-676-140	Sequence 140, App
32	56	3.1	416	10	US-09-960-352-4584	Sequence 4584, App
33	55.8	3.0	9539	9	US-10-239-676-52	Sequence 52, App1
34	55.6	3.0	960	9	US-10-198-846-6381	Sequence 6381, App1
35	55.4	3.0	377	10	US-09-960-352-7419	Sequence 7419, App
36	55.2	3.0	17848	9	US-10-239-676-28	Sequence 28, App1
37	55	3.0	480	10	US-09-960-352-5301	Sequence 5301, App
38	55	3.0	5689	9	US-09-938-842A-5191	Sequence 5191, App
39	55	3.0	2000	9	US-10-239-676-89	Sequence 89, App1
40	55	3.0	6053	9	US-10-239-676-76	Sequence 76, App1
41	55	3.0	11836	9	US-10-239-676-102	Sequence 102, App
42	55	3.0	14649	9	US-10-239-676-122	Sequence 122, App
43	54.8	3.0	6298	9	US-10-239-676-64	Sequence 64, App1
44	54.8	3.0	53332	9	US-10-224-562-3	Sequence 3, App1
45	54.8	3.0	53332	10	US-09-801-861-3	Sequence 3, App1

## ALIGNMENTS

RESULT 1  
US-10-076-622-208

Sequence 208, Application US/10076622  
Publication No. US20030023036A1

GENERAL INFORMATION:

APPLICANT: Houghton, Raymond L.

APPLICANT: Sleath, Paul R.

APPLICANT: Persing, David H.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.470C11

CURRENT APPLICATION NUMBER: US/10/076,622

NUMBER OF SEQ ID NOS: 627

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 208

LENGTH: 381

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

LOCATION: 10, 27, 37, 46, 75, 95, 102, 137, 143, 202, 234, 278, 310,

LOCATION: 351

OTHER INFORMATION: n = A,T,C or G

US-10-076-622-208

Query Match  
Best Local Similarity 95.3%; Pred. No. 8.7e-60;  
Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY	949	AGAGAGACTTGGCAGTCGAGTCTTCTTCCATCCACCCCTACACTTACATATTA	1008
DB	1	AGAGAGAGATNTTGGCAGTCGAGTCTTCTTCCATCCACCCCTACACTTACATATTA	60
QY	1009	CTTACTGCTTTGTTAAAGCAAGATATACCTTACTTGGCTTACTTGGCCTT	1068
DB	61	CTTACTGCTTTGTTAAAGCAAGATATACCTTACTTGGCTTACTTGGCCTT	120
QY	1069	TAGCTACTAATTAAGTTGATATAGCATTTATATATATCTGAGTCATTCATGTAT	1128
DB	121	TAGCTACTAATTAAGTTGATATAGCATTTATATATATCTGAGTCATTCATGTAT	180

QY 1129 CTCATCTTTGATGATTTTTCACAACTAAGATCATGATGTTTTCCTCAGAGTCC 1188  
 |||||||  
 DB 181 CTCATCTTTGATGATTTTTCACAACTAAGATCATGATGTTTTCCTCAGAGTCC 240  
 QY 1189 ATTAATCATTTATTTCTTACTTCTCCCTGTTGTAACATTTAGAACTGGATTT 1248  
 |||||||  
 DB 241 ATTAATCATTTATTTCTTACTTCTCCCTGTTGTAACATTTAGAACTGGATTT 300  
 QY 1249 GGAACCCCAATTTTGGAAAAACAGATTCATGATGAAAAATGCAATTCATATTCG 1308  
 |||||||  
 DB 301 GGAACCCCAATTTTGGAAAAACAGATTCATGATGAAAAATGCAATTCATATTCG 360  
 QY 1309 TTTTGAAGAGATGTGGCAT 1329  
 |||||||  
 DB 361 TTTTGAAGAGATGTGGCAT 381

## RESULT 2

US-09-604-287A-208  
 ; Sequence 208, Application US/09604287A  
 ; Patent No. US20020064872A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jiang, Yugu  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Xu, Jianshun  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Hepler, William T.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; FILE REFERENCE: 210121.470C7  
 ; CURRENT APPLICATION NUMBER: US/09/604,287A  
 ; CURRENT FILING DATE: 2000-06-22  
 ; NUMBER OF SEQ ID NOS: 489  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 208  
 ; LENGTH: 381  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(381)  
 ; OTHER INFORMATION: n-A,T,C or G  
 ; US-09-604-287A-208

Query Match 19.7%; Score 360.6; DB 10; Length 381;  
 Best Local Similarity 95.3%; Pred. No. 8.7e-60;  
 Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 949 AGAGAGACTTTTGGCATGCGCTGAGTCTTTCCTATCCCACTTAACATATTA 1008  
 |||||||  
 DB 1 AGAGAGATNTTGGCATGCGCTGAGTCTTTCCTATNCACCTTACATTAACATATTA 60  
 QY 1009 CTTAGTCTGCTTTGTTAAAGCAAGATTAACCTTAAGTCTTACTCTTGCCCTT 1068  
 |||||||  
 DB 61 CTTAGTCTGCTTTGTTAAAGCAAGATTAACCTTAACTTGKCTTACTCTTGCCCTT 120  
 QY 1069 TAGCTAACTAATAAGTTGATATAGGCATTAATATATATCTGATCATGGTAT 1128  
 |||||||  
 DB 121 TAGCTAACTAATAAGTTGATATAGGCATTAATATATATCTGATCATGGTAT 180  
 QY 1129 CTCATCTTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCAGAGTCC 1188  
 |||||||  
 DB 181 CTCATCTTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCAGAGTCC 240  
 QY 1189 ATTAATCATTTATTTCTTACTTCTCCCTGTTGTAACATTTAGAACTGGATTT 1248  
 |||||||  
 DB 241 ATTAATCATTTATTTCTTACTTCTCCCTGTTGTAACATTTAGAACTGGATTT 300  
 QY 1249 GGAACCCCAATTTTGGAAAAACAGATTCATGATGAAAAATGCAATTCATATTCG 1308  
 |||||||  
 DB 301 GGAACCCCAATTTTGGAAAAACAGATTCATGATGAAAAATGCAATTCATATTCG 360

QY 1309 TTTTGAAGAGATGTGGCAT 1329  
 |||||||  
 DB 361 TTTTGAAGAGATGTGGCAT 381

## RESULT 3

US-09-339-338-208  
 ; Sequence 208, Application US/09339338A  
 ; Patent No. US20020102602A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yugu, Jiang  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Xu, Jianshun  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
 ; FILE REFERENCE: 210121.470C2  
 ; CURRENT APPLICATION NUMBER: US/09/339,338A  
 ; CURRENT FILING DATE: 1999-06-23  
 ; NUMBER OF SEQ ID NOS: 315  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 208  
 ; LENGTH: 381  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(381)  
 ; OTHER INFORMATION: n-A,T,C or G  
 ; US-09-339-338-208

Query Match 19.7%; Score 360.6; DB 10; Length 381;  
 Best Local Similarity 95.3%; Pred. No. 8.7e-60;  
 Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 949 AGAGAGACTTTTGGCATGCGCTGAGTCTTTCCTATCCCACTTAACATATTA 1008  
 |||||||  
 DB 1 AGAGAGATNTTGGCATGCGCTGAGTCTTTCCTATNCACCTTACATTAACATATTA 60  
 QY 1009 CTTAGTCTGCTTTGTTAAAGCAAGATTAACCTTAAGTCTTACTCTTGCCCTT 1068  
 |||||||  
 DB 61 CTTAGTCTGCTTTGTTAAAGCAAGATTAACCTTAACTTGKCTTACTCTTGCCCTT 120  
 QY 1069 TAGCTAACTAATAAGTTGATATAGGCATTAATATATATCTGATCATGGTAT 1128  
 |||||||  
 DB 121 TAGCTAACTAATAAGTTGATATAGGCATTAATATATATCTGATCATGGTAT 180  
 QY 1129 CTCATCTTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCAGAGTCC 1188  
 |||||||  
 DB 181 CTCATCTTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCAGAGTCC 240  
 QY 1189 ATTAATCATTTATTTCTTACTTCTCCCTGTTGTAACATTTAGAACTGGATTT 1248  
 |||||||  
 DB 241 ATTAATCATTTATTTCTTACTTCTCCCTGTTGTAACATTTAGAACTGGATTT 300  
 QY 1249 GGAACCCCAATTTTGGAAAAACAGATTCATGATGAAAAATGCAATTCATATTCG 1308  
 |||||||  
 DB 301 GGAACCCCAATTTTGGAAAAACAGATTCATGATGAAAAATGCAATTCATATTCG 360  
 QY 1309 TTTTGAAGAGATGTGGCAT 1329  
 |||||||  
 DB 361 TTTTGAAGAGATGTGGCAT 381

## RESULT 4

US-10-007-805-208  
 ; Sequence 208, Application US/10007805  
 ; Patent No. US20020150581A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jiang, Yugu  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Xu, Jianshun



LENGTH: 5689  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 NAME/KEY: unsure  
 LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)  
 US-10-239-676-90

Query Match 3.8%; Score 70; DB 9; Length 5689;  
 Best Local Similarity 47.7%; Pred. No. 0.00094;  
 Matches 285; Conservative 0; Mismatches 286; Indels 4; Gaps 2;

QY 1274 TTCAATGTCATGAAGAAAGAACTCCATATTCGTTTGAAGAAAGTGGCCATTAT 1333  
 DB 1290 TTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1349  
 QY 1334 ACAGTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1393  
 DB 1350 ATATTTATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1409  
 QY 1394 GTTCGTGCAAGCTATACCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1453  
 DB 1410 TATATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 1469  
 QY 1454 CAGTTTAACTCAATGAGAG-TCAGAAATTTCTATGATTTGAGAAATGTTTAAATCAAT 1512  
 DB 1470 TTAATTTATTCGAGATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1529  
 QY 1513 CTATTAATCTGTAATTTCTAAGAGCTTAATTTGCTTTGCTGTAATGCTATATTG 1572  
 DB 1530 TATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 1589  
 QY 1573 AATTTGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 1632  
 DB 1590 ATTAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 1649  
 QY 1633 GATGTTAAATCAATAGATGATCTGTTTACATTTGTCATTAATTAATTAATTAATTTG 1692  
 DB 1650 TTAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 1709  
 QY 1693 ATTTCATTTGTAATTTGGCCTTAATTAATTTGTAATTTGTAATTTGTAATTTGTAAT 1752  
 DB 1710 ATT- -TTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1766  
 QY 1753 CTGTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 1812  
 DB 1767 TTAATTAATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1826  
 QY 1813 AAATTAATAATTTTA 1827  
 DB 1827 TTAATTTAAATTAATA 1841

RESULT 7  
 US-10-239-676-187  
 ; Sequence 187, Application US/10239676  
 ; Publication No. US20030082609A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OLEK, Alexander  
 ; APPLICANT: PIEPENBROCK, Christian  
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
 ; FILE REFERENCE: 5013.1003  
 ; CURRENT APPLICATION NUMBER: US/10/239, 676  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
 ; DE 10019058.8  
 ; DE 10019173.8  
 ; DE 10032529.7  
 ; DE 10043826.1  
 ; PRIOR FILING DATE: 2001-04-06

2000-04-06  
 ; 2000-04-07  
 ; 2000-06-30  
 ; 2000-09-01  
 ; NUMBER OF SEQ ID NOS: 228  
 ; SEQ ID NO 187  
 ; LENGTH: 11047  
 ; TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 US-10-239-676-187

Query Match 3.7%; Score 67.8; DB 9; Length 11047;  
 Best Local Similarity 43.5%; Pred. No. 0.0033;  
 Matches 306; Conservative 0; Mismatches 397; Indels 0; Gaps 0;

QY 1087 TGATATAGCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1146  
 DB 9873 TGATATAGCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTG 9932  
 QY 1147 TTTCAAACTAAGATCTATGATAGTTTTCACAGAGTCCATTAATTAATTAATTTCC 1206  
 DB 9933 TAAATGGGTGATTAATTTTATTTGATTAATTTGTAATTTTATTTTAAATTTTA 9992  
 QY 1207 TTTACTTCTCAGCTCTGTTGAACAATTAAGAACTGAAATTTGGGAACCAATTTTGGAA 1266  
 DB 9993 ATATTTATTTGGAAATTAAGAAATTTTAAATTTTAAATTTTAAATTAAGAAATTTTAA 10052  
 QY 1267 AACCAGATTCATGATCAAGAAATGAAGAACTCCATTAATTTGTTGAAAGATGGC 1326  
 DB 10053 GTGTGTAGGAATTAAGAAATTTGAAGAAATTTGTAATTTTGTAGTTAGGTT 10112  
 QY 1327 CATTAATCAAGTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTG 1386  
 DB 10113 TATTAATATTTATTAATTAATTTGTTTATTAATTAATTTTATTAATTTTATTTTAA 10172  
 QY 1387 ACAAGAGTTCGTGAGCAAGCTATACCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1446  
 DB 10173 GATGATTTTAAAGTAATTTAAGATTAATTAATTTTAAATTTTAAATTTGGAATTAAT 10232  
 QY 1447 GTTTGACGTTTAACTCAATGAGATCAAGATTAATTAATTAATTAATTAATTAATTTTAA 1506  
 DB 10233 AGTTGATTAATTTTGGTTTTTTTTTAAATTTGTAATTTGTAATTTGTAATTTGTAAT 10292  
 QY 1507 ATCAATCTAATTAATCTGTAATTTCTAAGAGCTTAATTTGCTTTGCTGTAATGTA 1566  
 DB 10293 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 10352  
 QY 1567 TATTTGAATTTGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 1626  
 DB 10353 GTTTAGTTAGTTACCTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 10412  
 QY 1627 ATATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAATTT 1686  
 DB 10413 AAATTAAGAGAAAGATTTAATTAATTTTGTTAATTAATTTTGTTTTAAATTTTAAATTT 10472  
 QY 1687 GTTTAATTTCAATTTGTAATTTGGCCTTAATTAATTTGTAATTTGTAATTTTAAATTTTCTCT 1746  
 DB 10473 TTTAAGATTAATTTGATTTGTAATTTAATTTAATTTAATTTTAAATTAATTTTAAATTTTAA 10532  
 QY 1747 CTGTTACTGTAATAATTAATTAATTAATTAATTAATTAATTAATTTTCTT 1789  
 DB 10533 TAAAGGTTTGTAGATTTATTTAGTTAATTTAATTTAATTTTAAATTTTAAATTTTAA 10575

RESULT 8  
 US-10-239-676-164  
 ; Sequence 164, Application US/10239676  
 ; Publication No. US20030082609A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OLEK, Alexander  
 ; APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIORITY FILING DATE: 2002-09-24  
PRIORITY APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIORITY FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 164  
LENGTH: 6030  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
NAME/KEY: unsure  
LOCATION: (3455)  
US-10-239-676-164

Query Match 3.7%; Score 67.6; DB 9; Length 6030;  
Best Local Similarity 48.4%; Pred. No. 0.0028;  
Matches 22; Conservative 0; Mismatches 230; Indels 7; Gaps 1;

QY 1320 ATGTGGCCATTTATTCAGTAATTTATTTATAGACTTGGCTGTCGACATTAATAGTAT 1379  
DB 5432 ATATGTAATTTATTTATATANAATAATTAATTAATTAATTAATTAATTAATTAAT 1491  
QY 1380 ATTTGGACAGAGTCTGTCGACCAAGCTTACCTAATTAATTAATTAATTAATTAATTA 1439  
DB 5492 TTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1551  
QY 1440 TATGAGTCTTTGACAGTTTAACTCAATGAGATCAGAAATTTATTTATTTATTTATTTAT 1499  
DB 5552 GGTGTATGATGAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1561  
QY 1500 GTTAAATATCAATCTAATAAT-----CTTGAATTTCTAAGAGCTTATTTGTTCTTT 1552  
DB 5612 GTTTTATTAATAAATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1561  
QY 1553 TGGCTGACAGATTAATTTGAATTTGTTGAATTAATTAATTTCTCATTTGAAAAATAT 1612  
DB 5672 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1612  
QY 1613 TATATCCCAAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1672  
DB 5732 TGTAGATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1672  
QY 1673 ATGATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1732  
DB 5792 ATTAATGTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1732  
QY 1733 ATTTTATTTCTCTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1771  
DB 5852 ATTTATTTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1771

RESULT 9  
US-10-239-676-197  
Sequence 197, Application US/10239676  
Publication No. US20030082609A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIORITY FILING DATE: 2002-09-24  
PRIORITY APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIORITY FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 197  
LENGTH: 7823  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-197

Query Match 3.6%; Score 65.4; DB 9; Length 7823;  
Best Local Similarity 45.3%; Pred. No. 0.0081;  
Matches 237; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY 1310 TTTTGAAGATGTCGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1369  
DB 2153 TTTTGAAGATTTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1429  
QY 1370 TATATGATTTATTTGACAGAGTCTGTCGACCAAGCTTACCTAATTAATTAATTAATTA 1429  
DB 2213 GTTAAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1489  
QY 1430 AAACATATATATGATGTTTGTACAGTTTAACTCAATGAGATCAGAAATTTATTTATTTATTT 1489  
DB 2273 AATGATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1549  
QY 1490 TTTGAGAAATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1549  
DB 2333 TTTGAGAAATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1549  
QY 1550 TTTTGGCTGAATGATTAATTTGATTTGTTGAATTAATTAATTTCTCATTTGAAAAAT 1609  
DB 2393 TATTTAAGAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1609  
QY 1610 AATTAATGCAAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1669  
DB 2453 AATGATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1669  
QY 1670 CATATGAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1729  
DB 2513 TAAATGTAATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1729  
QY 1730 CTAATTTATTTCTCTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1789  
DB 2573 TTTTAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1789  
QY 1790 CAGAAGATTTCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1832  
DB 2633 TAAATATGTTTAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1832

RESULT 10  
US-10-094-240-10  
Sequence 10, Application US/10094240  
Publication No. US20030082637A1  
GENERAL INFORMATION:  
APPLICANT: ZWIBEL, LAURENCE J.  
TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF  
FILE REFERENCE: N8289  
CURRENT APPLICATION NUMBER: US/10/094,240  
CURRENT FILING DATE: 2001-03-08

```

RESULT 12
US-10-239-676-18
/ Sequence 18, Application US/10239676
/ Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIORITY APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIORITY FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 18
LENGTH: 5979
TYPE: DNA
ORGANISM: Artificial Sequence
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-18

```







us-09-935-703-2.rnpb

Page 9

Search completed: June 11, 2003, 13:40:23  
Job time : 265 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 09:32:37 ; Search time 114 Seconds  
(without alignments)  
4928.349 Million cell updates/sec

Title: US-09-935-703-2

Perfect score: 1832  
Sequence: 1 ggcacgaggtatgatcatcat.....aaaataataattactctgt 1832

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCFUS.COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	68.8	3.8	19124	2	US-08-487-826B-13
C 2	67.2	3.7	837	4	US-08-998-416-288
C 3	65.2	3.6	20674	4	US-09-641-638-651
C 4	58.2	3.2	19124	2	US-08-487-826B-13
C 5	56.8	3.1	636	4	US-08-998-416-137
C 6	55	3.0	1634	4	US-08-913-842-20
C 7	55	3.0	1875	3	US-08-913-842-1
C 8	54.8	3.0	615	4	US-08-998-416-186
C 9	54.4	3.0	658	4	US-08-998-416-595
C 10	52.8	2.9	1511	1	US-07-991-867B-8
C 11	52.8	2.9	1511	1	US-08-107-755A-8
C 12	52.8	2.9	1511	2	US-08-544-332-8
C 13	52.8	2.9	1511	4	US-09-370-861A-8
C 14	51.8	2.8	5852	1	US-07-867-106-2
C 15	51.6	2.8	636	4	US-08-998-416-1137
C 16	51.6	2.8	837	4	US-08-998-416-288
C 17	51.2	2.8	2110	4	US-09-419-459-1
C 18	51.2	2.8	5923	4	US-09-064-922-3
C 19	50.8	2.8	660	1	US-07-991-867B-32
C 20	50.8	2.8	660	2	US-08-107-755A-32
C 21	50.8	2.8	660	2	US-08-544-332-32
C 22	50.8	2.8	660	4	US-09-370-861A-32
C 23	50.8	2.8	4810	3	US-08-852-629-11
C 24	50.8	2.8	4838	3	US-08-852-629-15
C 25	50.6	2.8	688	4	US-08-998-416-972
C 26	50.6	2.8	1422	1	US-08-319-704-5
C 27	50.4	2.8	7218	1	US-08-232-463-14

C 28	50.2	2.7	2430	4	US-08-845-258-3	Sequence 3, Appl1
C 29	50.2	2.7	2430	4	US-08-845-258-40	Sequence 40, Appl1
C 30	50.2	2.7	2430	4	US-08-990-571-3	Sequence 3, Appl1
C 31	50.2	2.7	2430	4	US-08-990-571-40	Sequence 40, Appl1
C 32	50.2	2.7	2430	4	US-08-723-142A-3	Sequence 3, Appl1
C 33	50.2	2.7	2430	4	US-08-723-142A-40	Sequence 40, Appl1
C 34	50.2	2.7	2430	4	US-09-528-784A-3	Sequence 3, Appl1
C 35	50.2	2.7	2430	4	US-09-528-784A-40	Sequence 40, Appl1
C 36	50	2.7	711	4	US-08-998-416-766	Sequence 766, App
C 37	50	2.7	1678	1	US-08-261-677-10	Sequence 10, Appl1
C 38	50	2.7	1678	1	US-08-384-556A-6	Sequence 6, Appl1
C 39	50	2.7	1678	2	US-08-331-355A-10	Sequence 10, Appl1
C 40	50	2.7	1678	5	PCT-US94-12364-10	Sequence 10, Appl1
C 41	50	2.7	1678	5	PCT-US95-07753-6	Sequence 6, Appl1
C 42	50	2.7	2327	3	US-09-157-077-1	Sequence 1, Appl1
C 43	50	2.7	2796	1	US-08-261-677-8	Sequence 8, Appl1
C 44	50	2.7	2796	1	US-08-384-556A-4	Sequence 4, Appl1
C 45	50	2.7	2796	2	US-08-331-355A-8	Sequence 8, Appl1

## ALIGNMENTS

RESULT 1  
US-08-487-826B-13/C  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chluis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Knobe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487, 826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121, 001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13  
Query Match 3.88; Score 68.8; DB 2; Length 19124;  
Best Local Similarity 48.5%; Pred. No. 1.1e-06;

	Matches	190; Conservative	0; Mismatches	202, Indels	0; Gaps	0;
QY	1435	ATGATATATGAGTGTGGACAGTTTAACATCAATGAGATCAGAAATTTCTAGTANTGG				1494
Db	16000	ATATATAATATATATAGTATATATGATGATGAATTAATTTATTTTATTTTATTTAA				15944
QY	1495	AAATGTTTATATATCATCATATATAATCTGAAATTTCTAAGAGGCTATTTGTTCTTGA				1554
Db	15940	TTAAATTTTTTTATTTTTTTTTTTTATTCATTAATTTTTTTAAATTTTTTTATTAATTA				15888
QY	1555	GCTGAATGACTAATTTTGAATTTGGTGAATATATTAATTAATTCATGTGAAAAATATTA				1614
Db	15880	AAATTTTTTATTTATATATTTTTTTTTTATTTAAATTAATTTTTTATTTATTTATTTTAA				15822
QY	1615	TATGCCAAAAATATTTTGAATGTTAAATCAAAAGATGATCTGTTACATGTCATAT				1674
Db	15820	TATTTAAATATAATTTTTTTTTTTATTTATGTAATAATTTTTTTTTTAACTTTTATATAT				15761
QY	1675	GAATTAATTAATCTGTGTTAATTCATTTTGATTAATTTGGCCTTAATTTGTATCTTAAT				1734
Db	15760	TTTTTATTTTATAGAAATATATTTTATTTTATTTAATATTTTTTTCTTTTTTTTTGTTTT				15700
QY	1735	TTTATTTTCCTCTGTTACTGTAATAATATAGCTAATAGTAAACAATTTCTTCAGAA				1794
Db	15700	TATGATATATATATTTTTTTTTTTTTTAAAGTTTTTTTTTTCTGTTGTTTATATTTT				15644
QY	1795	GAATTCATGCTATTAATTAATAAAATATAT				1826
Db	15640	TTATTAATCATTTTTTTTTTTTATATATAAAATTTTT				15609

RESULT 2  
 US-08-998-416-288  
 Sequence 288: Application US/08998416  
 Patent No. 6239264  
 GENERAL INFORMATION:  
 APPLICANT: Philippsen, Peter  
 APPLICANT: Pohlmann, Rainer  
 APPLICANT: Steiner, Sabine  
 APPLICANT: Mohr, Christine  
 APPLICANT: Wendland, Jurgen  
 APPLICANT: Knechtle, Philipp  
 APPLICANT: Redischung, Corinne  
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII  
 NUMBER OF INVENTION: AND USFS THEREOF  
 NUMBER OF SEQUENCES: 1152  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 6239264artis Corporation  
 STREET: 3054 Cornwallis Road  
 CITY: Research Triangle Park  
 STATE: No. 6239264th Carolina  
 COUNTRY: USA  
 ZIP: 27709  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/998,416  
 FILING DATE: 24-DEC-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 0016/97  
 FILING DATE: 31-DEC-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Melgs, J. Timothy  
 REGISTRATION NUMBER: 38, 241  
 REFERENCE/DOCKET NUMBER: PE/5-30306/A/GCG1976  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 288:

```

?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 837 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      ORIGINAL SOURCE:
?      ORGANISM: PAC1241RP
?      IS-08-998-416-288

```

Query Match	Score	DB	Length
Best Local Similarity	3.78	4	837
Best Local	67.2	4	837
Best Local	47.68	4	837
Best Local	10.06	4	837

Matches 198; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

1416 AATTATAAGCTATAAACATAGATATGAGTGTGTGACAGTTTAACTCAATGGAGATCA 1475

190 AATAATATTTAATATGAATACTATTAGTCTATGTCCAATTTTAAATTAGTTATTTAAA 249

1476 GAATATTCCTATGTATTGAGAAATGTTAATATCAATCTATAATCTTGAA TTC TAAGA 1535

250 TATTATTAGATATTATTCTTTAATAAATTATTAAATAGATTATCAATAATTATA 309

1536 GGCCTATTCTCTTTGGCTGAATGAGTATATTGCAATTGGTTGAATTAATAATT 1595

310 TATTATTATTAAATGTTTATTAAATTAATTTTATTATTATAAAGATTTAATTAT 369

1596 CTCATTGTAATAATTATATGCCAAATAATTTGATGTTAAATCAATAGATGATT 1655

370 TTAATATTTGTAATTTATTTTATTTATTAATACTATTTTATTAATAATATTATGTGGAT 429

1656 CTGTTTACATTGTTTCATATGAATAATCTGTGTTAATTTCATTTTGATAATTGGCCTT 1715

430 TTATATTATTAACTTTTATAGGAATTATTATAAAATTAAATTTTAACTTTAAATTTCT 489

1716 TAATAATTGTAATCTAATTTTATTTCTCTGTACTGTAATAATAGCTATAATGT 1775

490 TATTATTAATTTTATATTTAATAAATTATATTTCAATTTATTTATTTATTTA 549

1776 ATACCAATTTCTTCAGAAGAAATCTATGCTATTATTAAATAAATATTTACTGT 1831

550 ATTAAATTATTTAATTATTTTATCATTTAATTAAATTAAATAAT 605

## 2. THEOREM

S-09-641-638-651  
SECURITY 651 April 1994  
SECURITY 651 April 1994

Patent No. 6432648

APPLICANT: Blumenfeld, Marta  
APPLICANT: Bouquelerot, Lydia

APPLICANT: Chumakov, Ilya  
APPLICANT: Cohen, Arnold

[illegible]

FILE REFERENCE: GENSET.051CPI  
CLIPBENT APPLICATION NUMBER: ITS/08/641 639

CURRENT FILING DATE: 2000-08-16  
PRIOR APPLICATION NUMBER: IIS 09/502 330

PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: IIS 60/133 200

PRIOR FILING DATE: 1999-05-07  
PRIOR APPLICATION NUMBER: US 09/275,267

PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: IIS 60/119.917

PRIOR FILING DATE: 1999-02-12  
NUMBER OF SEQ ID NOS: 1304

SOFTWARE: patent.pm  
SEQ ID NO 651

LENGTH: 206/4  
TYPE: DNA

ORGANISM: *Homo sapiens*  
FEATURE:

NAME/KEY: misc\_feature

---

```
LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5996..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1837
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
```

NAME/KEY: allele  
LOCATION: 8777  
OTHER INFORMATION: 10-349-216 : deletion of CTG  
NAME/KEY: allele  
LOCATION: 8785  
OTHER INFORMATION: 10-349-224 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 8926  
OTHER INFORMATION: 10-349-368 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12171  
OTHER INFORMATION: 10-350-72 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12429  
OTHER INFORMATION: 10-350-332 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13341  
OTHER INFORMATION: 10-507-170 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 13492  
OTHER INFORMATION: 10-507-321 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 13524  
OTHER INFORMATION: 10-507-353 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13535

Query Match 3.6%; Score 65.2; DB 4; Length 20674;  
Best Local Similarity 49.0%; Pred. No. 7.6e-06;  
Matches 236; Conservative 0; Mismatches 238; Indels 8; Gaps 2;

1344 TATTATAGCATTGGCTGCTACAAATTAATGATATTTTGCAAGAGACTGCTGTA 1403  
11085 TATTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11144  
1404 CAAGCTATACCTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1463  
11145 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11204  
1464 CAATGAGATCAGAAATTTCTATGATTTGAGAAATGTTTAAATCAATCTAATAATCTT 1523  
11205 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11264  
1524 GAATTTCTAAGAGCTTATTTGCTTTTGGCTGAATGAGTAAATTTGAATGCTTGA 1583  
11265 TATTATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11324  
1584 TATTATTAATTTCTCATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 1643  
11325 TTAATTAATTAATTAATTAATTAATTAATTAATTTT---AATATTAATTAATTAATTT 11380  
1644 AATATGATGATTTCTGTTACATTTGTCATTAATTAATTAATTAATTAATTAATTTTG 1703  
11381 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTA 11440  
1704 AATATGAGCTTAAAT---TTGATCTCAATTAATTTTCTGCTGTTACTGTA 1759  
11441 AATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAA 11500  
1760 AATATGATGATTTGTTAATTAATTAATTTCTCAGAAATGATTAATTAATTAATTAATTA 1819  
11501 TTAATTAATTAATTAATTAATTAATTTAATGCTAAGAAATTAATTTAATTAATTA 11560  
1820 AA 1821  
11561 GA 11562

RESULT 4  
US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhuan  
APPLICANT: Williams, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-0176  
TELEFAX: (619) 235-8550

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 3.2%; Score 58.2; DB 2; Length 19124;  
Best Local Similarity 46.4%; Pred. No. 0.0003;  
Matches 189; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

1420 AATAGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1479  
407 AA 466  
1480 ATTCTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1539  
467 GTTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 526  
1540 TATTGTTCTTTGGCTCAATGAGTAAATTTGTTGATTAATTAATTAATTAATTTCTCA 1599  
527 AATTAACATAGAAAGATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 586  
1600 TTGTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1659  
587 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 646  
647 TTTCTTTTATCAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 706  
1720 ATTGATCTCAATTTAATTTTCTGCTGTTAATTAATTAATTAATTAATTAATTAATTAAT 1779  
707 CGTTATATACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 766  
1780 CAATTTCTTCAAGAAATTTCTATGCTAATTAATTAATTAATTAATTT 1826





TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 3.0%; Score 54.8; DB 4; Length 615;  
Best Local Similarity 47.4%; Pred. No. 0.00065;  
Matches 197; Conservative 0; Mismatches 217; Indels 2; Gaps 1;

QY 1416 AATTAATGCTATTAACATAGATGAGTGTGTTGACAGTTAACTGATGAGATCA 1475  
DB 190 AATAATATTAATGAACTATGCTATGCTATGCTAATTTAATAGTATATAAA 249  
QY 1476 GAATATTCATGCTATGAGAAATGTTAATATCACTATTAATCTTGCAATTCAGACA 1535  
DB 250 TATTAATGATATTAATTTATTTCTTTAATTAATTAATGATATCATATTAATA 309  
QY 1536 GCGTATTTGTTGCTTTGGCTGATGAGTATATTTGATTTGGTGAATTAATAAT 1595  
DB 310 TATTTATTTATTAATGTTTATTAATAATATATTTATTTATTAATTAATTTAT 369  
QY 1596 CTCATTTGTAATAATATATATATGCAAAATATATTTGATGTTAAATCAATGATGAT 1655  
DB 370 TTAATATTTGTAATAATATATTTATTTATTAATATCTATTTTATTAATAATTT -ATGTG 427  
QY 1656 CTGTTACATGCTATGATGAAATATCTGCTGTTAATTTCAATTTGATTTGGCTT 1715  
DB 428 ATTAATTAATTAATTAATCTTTTATTAAGAAATTAATTAATTAATTAATTAATTC 487  
QY 1716 TAAATTTGTAATCTAATTTATTTCTGCTGATGTAATAATAGCTATATCT 1775  
DB 488 TATATTAATTTATTTATTTATTTATTAATTAATTAATTTATTTATTTATTTATTTA 547  
QY 1776 ATAAATTTCTTCTGAGAAATCTATGCTATTTATTAATAATTAATTTACTGT 1831  
DB 548 ATTAATTAATTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTAATAAT 603

## RESULT 9

US-08-998-416-595  
Sequence 595, Application US/08998416  
Patent No. 6239264

## GENERAL INFORMATION:

APPLICANT: Philippesen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Redischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwalls Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 595:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 658 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1408RP  
US-08-998-416-595

Query Match 3.0%; Score 54.4; DB 4; Length 658;  
Best Local Similarity 47.4%; Pred. No. 0.00082;  
Matches 163; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1488 TATGAGAAATGTTAATCAATCTAATAATCTGATTTCTAAGAGGCTATTTGT 1547  
DB 186 TATTAATCTTATCTGATTAATTAATTTATTTATTTATTTATTTATTTATTTAT 245  
QY 1548 TCTTTGGCTGAATGAGTATATTTGAATTTGTTGATTAATTAATTTCTCATTTGAAA 1607  
DB 246 TATTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 305  
QY 1608 ATATTTATATGCCAAATATTTATTTGATTTGATTAATCAATATGATCTGTTACATG 1667  
DB 306 TTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 365  
QY 1668 TTCAATGATTAATTAATCTGCTGTTAATTTCAATTTGATTAATTTGGCTTAATTTGAT 1727  
DB 366 GTTGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 425  
QY 1728 CTCATTTATTTATTTCTCTGCTGTTAATTAATTAATTAATTAATTAATTAATTTTC 1787  
DB 426 TTTTACTTATGATATATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTA 485  
QY 1788 TTCAAGAAATCTATGCTATTTAATAATTAATTAATTTACTGT 1831  
DB 486 TTCAATTAATGATTTGATTTGTTGACATATATATATGCTAT 529

## RESULT 10

US-07-991-867B-8/c  
Sequence 8, Application US/07991867B  
Patent No. 5476781

## GENERAL INFORMATION:

APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/991,867B



FILING DATE: 12-DEC-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: WO 92/14818  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA: US 07/827,685  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UFI14.C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Amsacta moorei entomopoxvirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (18..218)  
NAME/KEY: CDS  
LOCATION: complement (234..782)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 852..1511  
US-07-991-867B-8

Query Match 2.9% Score 52.8; DB 1; Length 1511;  
Best Local Similarity 49.5%; Pred. No. 0.0025;  
Matches 196; Conservative 0; Mismatches 192; Indels 8; Gaps 2;  
QY 1416 AATTAAAGCTATAAACAATAGATGAGTGTGTACAGTTAACTCAATGAGATCA 1475  
DB 1209 AATTAAAGCTATAAACAATAGATGAGTGTGTACAGTTAACTCAATGAGATCA 1475  
QY 1476 GAATATTCATGATGAGAAATGTTAATATCACTAATTAATCTTGAATTTCTAGA 1535  
DB 1149 TAATATTTACTTAATATGCTATTTTGTAAAGATATATCTAAATATGTTAATTTTA 1090  
QY 1536 GCGTATTTTGTCTTTGGCT--GAATGATATATTTGAATGTTGATTAATTAATA 1593  
DB 1089 ATTGTGTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1030  
QY 1594 TT-----CTCATTGTAATAAATAATTAATGACCAAAATAATATTTGATTAATCA 1647  
DB 1029 TTCTGGAATATATTAATAATATTAATTAATTAATTAATTAATTAATTAATTAAT 970  
QY 1648 AGATGATTCGTTTACATTTCTCATAGAAATTAATCTGTTAATTTCTATTGATTA 1707  
DB 969 TAATATTTTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 910  
QY 1708 TTGGCCTTAATTTGATCTCTAATTTTCTCTGTTAGCTAATAATTAATAGC 1767  
DB 909 TTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 850  
QY 1768 TATATGATATACAAATTTCTCAGAGAATTTCTAT 1803  
DB 849 TGATTAATTTATTTTTCATGATTAATTTT 814  
RESULT 11  
US-08-107-755A-8/c

Sequence 8, Application US/08107755A  
Patent No. 5721352  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5721352e1 Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/107,755A  
FILING DATE: 19-AUG-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,658  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UFI14.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Amsacta moorei entomopoxvirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (18..218)  
NAME/KEY: CDS  
LOCATION: complement (234..782)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 852..1511  
US-08-107-755A-8

Query Match 2.9% Score 52.8; DB 1; Length 1511;  
Best Local Similarity 49.5%; Pred. No. 0.0025;  
Matches 196; Conservative 0; Mismatches 192; Indels 8; Gaps 2;  
QY 1416 AATTAAAGCTATAAACAATAGATGAGTGTGTACAGTTAACTCAATGAGATCA 1475  
DB 1209 AATTAAAGCTATAAACAATAGATGAGTGTGTACAGTTAACTCAATGAGATCA 1475  
QY 1476 GAATATTCATGATGAGAAATGTTAATATCACTAATTAATCTTGAATTTCTAGA 1535  
DB 1149 TAATATTTACTTAATATGCTATTTTGTAAAGATATATCTAAATATGTTAATTTTA 1090  
QY 1536 GCGTATTTTGTCTTTGGCT--GAATGATATATTTGAATGTTGATTAATTAATA 1593  
DB 1089 ATTGTGTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1030

```

QY 1594 TT-----CTCATGTGTAATAATATATATGCGCAAAAATATATGATGTTAAATCAAT 1647
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1029 TTTCTGGAATATTTATTTAAATATATATATATATATATATATATATATATATATATAT 970
QY 1648 AGATGATTCCTGTTACATTTGTCATATGATATATATATATATATATATATATATATATAT 1707
    || || || || || || || || || || || || || || || || || || || || || ||
Db 969 TAACATAATTTTATATATATATATATATATATATATATATATATATATATATATATATAT 910
QY 1708 TTGGCCCTTAAATATTTGATCTATATATATATATATATATATATATATATATATATATAT 1767
    || || || || || || || || || || || || || || || || || || || || || ||
Db 909 TTAATAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 850
QY 1768 TATAATGATATACAAATTTCTTCAGACAAATTCAT 1803
Db 849 TGATAATTTTATATTTTCTTCATGATTAATTTT 814

RESULT 12
; Sequence 8, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5935777e1 Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE: US/08/544,332
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UF114.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:

```

```

; ORGANISM: Amsacta moorei entomopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18...218)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (234...782)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 852...1511
; US-08-544-332-8

Query Match      2.98; Score 52.8; DB 2; Length 1511;
Best Local Similarly 49.5%; Pred. No. 0.0025;
Matches 196; Conservative 0; Mismatches 192; Indels 8; Gaps 2;

QY 1416 AATTATAGCTATTAACATATAGATATAGAGTGTGTAGAGTTTAACTCATGAGATCA 1475
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1209 AATTATAGCTATTTATTTACATGATTCACAAATTTAAATTCATAGAAATGAGTACTA 1150
QY 1476 GAATATTCATGATTTGAGAAATGTTAAATCATCTATATATATATATATATATATATAT 1535
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1149 TAATATTACTATATTTGCTATTTTGTATTAAGATATATCTAAATATGTTATATTTTAA 1090
QY 1536 GCGTTATTTGTTCTTTGGCT--GAATGATATTTGAAATGTTGTTGAATATTAATA 1593
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1089 ATTTGTTATTAATTTAAATATATATATATATATATATATATATATATATATATATAT 1030
QY 1594 TT-----CTCATGTGTAATAATATATATATGCGCAAAAATATATGATGTTAAATCAAT 1647
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1029 TTTCTGGAATATTTATTTAAATATATATATATATATATATATATATATATATATATAT 970
QY 1648 AGATGATTCCTGTTACATTTGTCATATGATATATATATATATATATATATATATATATAT 1707
    || || || || || || || || || || || || || || || || || || || || || ||
Db 969 TAACATAATTTTATATATATATATATATATATATATATATATATATATATATATATATAT 910
QY 1708 TTGGCCCTTAAATATTTGATCTATATATATATATATATATATATATATATATATATATAT 1767
    || || || || || || || || || || || || || || || || || || || || || ||
Db 909 TTAATAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 850
QY 1768 TATAATGATATACAAATTTCTTCAGACAAATTCAT 1803
Db 849 TGATAATTTTATATTTTCTTCATGATTAATTTT 814

RESULT 13
; Sequence 8, Application US/09370861A
; Patent No. 6410221
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 6410221e1 Entomopoxvirus Expression System
; FILE REFERENCE: UF114.C4.D1
; CURRENT APPLICATION NUMBER: US/09/370,861A
; CURRENT FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 07/991,867
; PRIOR FILING DATE: 1992-12-07
; PRIOR APPLICATION NUMBER: US 08/107,755
; PRIOR FILING DATE: 1993-08-19
; PRIOR APPLICATION NUMBER: WO 92/14818
; PRIOR FILING DATE: 1992-02-12
; PRIOR APPLICATION NUMBER: US 07/827,685
; PRIOR FILING DATE: 1992-01-30
; PRIOR APPLICATION NUMBER: US 07/657,584
; PRIOR FILING DATE: 1991-02-19
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8
; LENGTH: 1511
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus

```



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8587  
INFORMATION FOR SEQ. ID NO: 1137:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1692RP  
US-08-998-416-1137

Query Match 2.8%; Score 51.6; DB 4; Length 636;  
Best Local Similarity 47.4%; Pred. No. 0.0036;

Matches 219; Conservative 0; Mismatches 239; Indels 4; Gaps 2;

QY 1364 TACAATTAACTGATTTTGGACAGAGTCTGGTGACAGCTATACCTAATTATA 1423  
DB 477 TAAATTAATTTAATAAATTTCTATATAAAGATTAAATATAATCAACATAATA 418  
QY 1424 GCTATAAACATATAGATGTTGTACAGTTTAACCAATGAGATCAGAAATATTC 1483  
DB 417 TTTATATAAATGATATTAATAAATAAATAATTAATTAATTAATAATTAATTAATC 358  
QY 1484 TATGATATGAGAAATGTTAATATCAATCTAATCTGAATTTCTAAGAGCTTAT 1543  
DB 357 TTTATATAAATGATATTAATAAATAAATAATTAATTAATTAATTAATTAATTAATG 298  
QY 1544 TTGTTCTTTGGCTGAATGAGTATAT--TGAATGTTGAATTAATTAATTAATTCAT 1600  
DB 297 ATAATCTAATTAATTAATTAATAAATAATATATATATATTAATTAATACTA 238  
QY 1601 TGTAAATATATATATGCCAAAATATATTTGATGTTAATCAATAGATGATTCGTT 1660  
DB 237 ATTTAAATTTGACATAGACTAAATAGTATTAATTAATTAATTAATTAATTAATA 178  
QY 1661 TACATTTGTCATGATATATATATCTGTTAATTTCAATTTGATTAATGGCTTAATA 1720  
DB 177 TAAATATTAATATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 118  
QY 1721 TTGTAATCTAATTAATTAATTTCTCTGTTACTGTAATTAATTAATGCTAATGTAATAC 1780  
DB 117 ATCAAAATATATTA--TTTATTAATAATGATATATATAGTTTAATAATTAATTAATTAATA 59  
QY 1781 AATTTCTGAGAGAAATCTATGCTATTAATTAATAAAT 1822  
DB 58 TATTAATAAATAAGTTTATTAATTAATTAATTAATTAATAAAT 17

Search completed: June 11, 2003, 11:46:36  
Job time: 117 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 07:31:33 ; Search time 430 Seconds  
(Without alignments)

9594.560 Million cell updates/sec

Title: US-09-935-703-2

Perfect score: 1832  
Sequence: 1 ggcacgaggtacatacatcat.....aaataaataattactgtg 1832

Scoring table: IDENTITY\_NGC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1832	100.0	1832	22	AAH43389
2	1688.6	92.2	3264	23	AAST0229
3	576.4	31.5	1914	24	AAAD24022
4	360.6	19.7	381	22	AAAF17150
5	360.6	19.7	381	22	AAAF17120
6	126.4	6.9	1785	20	AAV81744
7	126.4	6.9	1896	20	AAV81745
8	83.2	4.5	50000	24	ABL55644
9	79.4	4.3	5413	22	AA546694

10	78.6	4.3	1692	20	AAV81746
11	54.7	3.9	7333	22	AA546758
12	72	3.9	7333	24	ABL92318
13	72	3.9	7333	24	ABL34124
14	71.8	3.9	15732	22	AA545388
15	71.8	3.9	15732	24	ABL28233
16	71.4	3.9	7165	24	ABL70222
17	71.4	3.9	7165	24	ABL32751
18	71.4	3.9	7165	24	ABL31267
19	70.8	3.9	6118	24	ABL33031
20	70.4	3.8	11805	24	ABL33748
21	70	3.8	5689	22	AA545384
22	70	3.8	5689	22	AA546426
23	70	3.8	5689	24	ABL28226
24	69.2	3.8	5185	24	ABL33034
25	68.8	3.8	6365	24	ABL32124
26	68.8	3.8	8781	24	ABL33687
27	68.8	3.8	9155	24	ABL32462
28	68.8	3.8	19124	18	AA772882
29	68.8	3.8	19124	21	AA298287
30	68.6	3.7	11047	22	AA546664
31	67.8	3.7	11047	22	AA545479
32	67.8	3.7	11047	24	ABL33984
33	67.8	3.7	11047	24	ABL28401
34	67.6	3.7	6030	22	AA545456
35	67.6	3.7	6030	24	AA561348
36	67.6	3.7	6030	24	ABL28312
37	67.6	3.7	6292	22	AA546735
38	67.6	3.7	11155	24	ABL32604
39	67.2	3.7	9789	17	AA14852
40	67	3.7	9642	24	ABL33356
41	66.8	3.6	11996	24	ABL34492
42	66.4	3.6	700	22	AAH93026
43	66.2	3.6	6175	24	ABL33307
44	65.8	3.6	6106	22	AA546429
45	65.8	3.6	6106	24	ABL40031

#### ALIGNMENTS

RESULT 1	AAH43389	standard; CDNA; 1832 BP.
ID	AAH43389	
XX	AAH43389	
AC	15-NOV-2001	(first entry)
DT		
XX		
DE	PTPase cDNA clone HATBM23.	
XX		
KW	Protein tyrosine phosphatase; PTPase; neural disorder; preservative;	
KW	Alzheimer's disease; Parkinson's disease; immune system disorder;	
KW	rheumatoid arthritis; Grave's disease; muscular disorder; wound healing;	
KW	reproductive disorder; pulmonary disorder; cardiovascular disorder;	
KW	arrhythmia; infectious disease; viral hepatitis; multiple sclerosis;	
KW	hyperproliferative disorder; neoplasm; epithelial cell proliferation;	
KW	endocrine disorder; diabetes mellitus; transplantation; weight disorder;	
KW	hair loss; skin aging; sunburn; food additive; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	19..222
FT		/*tag= a
XX		/product= "PTPase"
PN	W0200166706-A1.	
XX		
PD	13-SEP-2001.	
XX		
PF	22-FEB-2001; 2001WO-US05496.	
XX		

PR 03-MAR-2000; 2000US-0186658.  
PR 16-MAR-2000; 2000US-0189881.  
XX  
PA (HMAN-) HUMAN GENOME SCI INC.

XX Shl Y, Ruben SM;  
XX MPI: 2001-565582/63.  
DR P-PSDB: AAB47496.

XX Novel isolated protein tyrosine phosphatase polypeptide useful for  
XX treating and preventing Alzheimer's disease, rheumatoid arthritis,  
XX Grave's disease, arrhythmias, neoplasms, multiple sclerosis and  
XX diabetes mellitus

PS Claim 1; Page 276-77; 286pp; English.

XX The sequences given in AAB43389-93 encode protein tyrosine phosphatase  
XX (PTPase) polypeptides. The PTPase polypeptides, or the cDNA encoding  
XX them, are useful for preventing, treating or ameliorating a medical  
XX condition in a mammalian subject. They are useful for diagnosing,  
XX preventing or treating neural disorders (e.g. Alzheimer's disease,  
XX Parkinson's disease), immune system disorders (e.g. rheumatoid  
XX arthritis, Grave's disease), muscular disorder, reproductive  
XX disorders, pulmonary disorders, cardiovascular disorders  
XX (e.g. arrhythmias), infectious diseases (e.g. viral hepatitis),  
XX hyperproliferative disorders (e.g. neoplasms), diseases at the cellular  
XX level (e.g. multiple sclerosis), endocrine disorders (e.g. diabetes  
XX mellitus), and wound healing and epithelial cell proliferation.  
XX They are also useful for preventing hair loss, to prevent skin aging  
XX due to sunburn, to maintain organs before transplantation, to modulate  
XX mammalian characteristics, to treat weight disorders, to change a  
XX mammal's mental or physical state, or as a food additive or  
XX preservative.

XX Sequence 1832 BP; 560 A; 299 C; 321 G; 652 T; 0 other;

Query Match 100.0%; Score 1832; DB 22; Length 1832;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACGAGGTATGATATCATGACGAAATTTATGCTTTAGAACTTAAGAAATCTGCTGT 60  
DB 1 GGCACGAGGTATGATATCATGACGAAATTTATGCTTTAGAACTTAAGAAATCTGCTGT 60  
QY 61 GAGTTAACTCTGGGAAATCAACCAAGCAAGAGAAAAAACAAGATCCGAGATATCTT 120  
DB 61 GAGTTAACTCTGGGAAATCAACCAAGCAAGAGAAAAAACAAGATCCGAGATATCTT 120  
QY 121 CCATTTCAACATCATGATATAGTGGCCCAATGAGAGAACAGTCTGGCATGTTCA 180  
DB 121 CCATTTCAACATCATGATATAGTGGCCCAATGAGAGAACAGTCTGGCATGTTCA 180  
QY 181 AACGAAGAGCACTATCATCTTTGTTACGATATTTGCTTGAACTTCTTGGAAACTTCT 240  
DB 181 AACGAAGAGCACTATCATCTTTGTTACGATATTTGCTTGAACTTCTTGGAAACTTCT 240  
QY 241 GACTTGGATTTAGAAAGACTTCTGCTCTCACTGAAATTAACAAGGGGTTTGA 300  
DB 241 GACTTGGATTTAGAAAGACTTCTGCTCTCACTGAAATTAACAAGGGGTTTGA 300  
QY 301 CCTCTCATAAAGAACATGTTTGCACACTGTGTAAGGGCTTTGCTATGCATACATCTGC 360  
DB 301 CCTCTCATAAAGAACATGTTTGCACACTGTGTAAGGGCTTTGCTATGCATACATCTGC 360  
QY 361 TTTCTGTTTATCAGTTATTTTCTTTCTTAAAGCTCCCTGAAGGGCAATATCATTTGG 420  
DB 361 TTTCTGTTTATCAGTTATTTTCTTTCTTAAAGCTCCCTGAAGGGCAATATCATTTGG 420  
QY 421 CTTGGGGTATCAGTTTCTTATGATCTGCTAGGCAATATCAAAATTAATCTCCAC 480  
DB 421 CTTGGGGTATCAGTTTCTTATGATCTGCTAGGCAATATCAAAATTAATCTCCAC 480

QY 481 ATTTCCAGTGAACAGATGTTTACATTAACGATTGCACTTGGCTATTGTTGAAGG 540  
DB 481 ATTTCCAGTGAACAGATGTTTACATTAACGATTGCACTTGGCTATTGTTGAAGG 540  
QY 541 ATTACAGAGCCCATTAAGGATTTAAATATATTCATTAAGATTTATTTGGAAAGTGG 600  
DB 541 ATTACAGAGCCCATTAAGGATTTAAATATATTCATTAAGATTTATTTGGAAAGTGG 600  
QY 601 CTGAGAGAGCTGAGATTTCCAGACTTTGTAAGTCTTATCTGGGAGAACATAGGC 660  
DB 601 CTGAGAGAGCTGAGATTTCCAGACTTTGTAAGTCTTATCTGGGAGAACATAGGC 660  
QY 661 CAATATATCATGACCTTTCCAGGCAATTTTAAAGACAGATGCTATTCATGTTCTTAGT 720  
DB 661 CAATATATCATGACCTTTCCAGGCAATTTTAAAGACAGATGCTATTCATGTTCTTAGT 720  
QY 721 AGACCTGCTACTTTTGTGTCATTTGAATTAACCAAGTTTAAAGAGCCAGTAGGGT 780  
DB 721 AGACCTGCTACTTTTGTGTCATTTGAATTAACCAAGTTTAAAGAGCCAGTAGGGT 780  
QY 781 GGACTTAATTTGACACAAATTTGGCTCATTTCTTACATTTTCACTGCTGCTCTCT 840  
DB 781 GGACTTAATTTGACACAAATTTGGCTCATTTCTTACATTTTCACTGCTGCTCTCT 840  
QY 841 ACAGCTGCTAGACCAAGACCTGTTGCTGGAGACATTTCAATGATAGGAGAGCTCT 900  
DB 841 ACAGCTGCTAGACCAAGACCTGTTGCTGGAGACATTTCAATGATAGGAGAGCTCT 900  
QY 901 CTCGGTGAACAGTCCAAACATTAATAGATTTTATGAAAGCCCAAGAGAGACTT 960  
DB 901 CTCGGTGAACAGTCCAAACATTAATAGATTTTATGAAAGCCCAAGAGAGACTT 960  
QY 961 TGCCATGCTGAGTCTTCTTCCATCCACCTTACACCTTAACATATTAATGATGCTT 1020  
DB 961 TGCCATGCTGAGTCTTCTTCCATCCACCTTACACCTTAACATATTAATGATGCTT 1020  
QY 1021 TGTAAAGCAAGATTAATACCTTTAACTGCTTACTCTTCTGCTTACTTACTTAA 1080  
DB 1021 TGTAAAGCAAGATTAATACCTTTAACTGCTTACTCTTCTGCTTACTTACTTAA 1080  
QY 1081 AAAGTTTATATGAGCATTTATATTAATCTGAGCATTCATGATCTCATCTTTG 1140  
DB 1081 AAAGTTTATATGAGCATTTATATTAATCTGAGCATTCATGATCTCATCTTTG 1140  
QY 1141 ATGATTTTCAAACTAAGATATGATAGTATTTTTCACAGAGTCCATTAATCATTT 1200  
DB 1141 ATGATTTTCAAACTAAGATATGATAGTATTTTTCACAGAGTCCATTAATCATTT 1200  
QY 1201 ATTCTCTTACTTCTACCTCTGTTGAACATTTTGAACATGATTTGGAAACCAAT 1260  
DB 1201 ATTCTCTTACTTCTACCTCTGTTGAACATTTTGAACATGATTTGGAAACCAAT 1260  
QY 1261 TTGGAACACGATTCATAGTATGATGAATTAAGAACTTCCATATCTGTTTGAAGA 1320  
DB 1261 TTGGAACACGATTCATAGTATGATGAATTAAGAACTTCCATATCTGTTTGAAGA 1320  
QY 1321 TGTGGCCATTTATACAGTAATTTTATATGAGACTTTGCTGCTAGCAATTAATAGTANA 1380  
DB 1321 TGTGGCCATTTATACAGTAATTTTATATGAGACTTTGCTGCTAGCAATTAATAGTANA 1380  
QY 1381 TTTTGGACAAGAGTCTGCTGACAACTATACCTTAATTAAGCTATTAACAAATAGAT 1440  
DB 1381 TTTTGGACAAGAGTCTGCTGACAACTATACCTTAATTAAGCTATTAACAAATAGAT 1440  
QY 1441 ATGAGTGTGTTAGAGTTTAACTCAATGAGAGATAGATATTTCTATGTTAGGAAGAT 1500  
DB 1441 ATGAGTGTGTTAGAGTTTAACTCAATGAGAGATAGATATTTCTATGTTAGGAAGAT 1500  
QY 1501 TTTATATCAATCTTAAATCTTGAATTTCTAAGAGGCTTATTTGCTTTGGCTGA 1560  
DB 1501 TTTATATCAATCTTAAATCTTGAATTTCTAAGAGGCTTATTTGCTTTGGCTGA 1560  
QY 1561 TGAGTATATTTGAATTTGTTGAATTAATTAATTTCTATGTTGAAGAAATTAATATGCT 1620



QY 1025 AAAAGCAAGTATTAACCTTAACTGCTCTACTCTTCCCTTTAGCTAACTAATAAG 1084  
 |||||||  
 Db 2458 AAAAGCAAGTATTAACCTTAACTGCTCTACTCTTCCCTTTAGCTAACTAATAAG 2517  
 |||||||  
 QY 1085 TTGTGATATGAGCATTTATATATATTTGAGCATTCATGATATCTCTCATGTTTGATGT 1144  
 |||||||  
 Db 2518 TTGTGATATGAGCATTTATATATATTTGAGCATTCATGATATCTCTCATGTTTGATGT 2577  
 |||||||  
 QY 1145 ATTTTCAACAGTATCATATATAG - TTTTTCACAGATTCATTAATCATTTATTT 1203  
 |||||||  
 Db 2578 ATTTTCAACAGTATCATATATAGTTTTCACAGATTCATTAATCATTTATTT 2637  
 |||||||  
 QY 1204 TCTTACTTCTCACTCTCTGTTGAACATTTAGAACTGGATTTGGAAACCAATTTTG 1263  
 |||||||  
 Db 2638 TCTTACTTCTCACTCTCTGTTGAACATTTAGAACTGGATTTGGAAACCAATTTTG 2697  
 |||||||  
 QY 1284 GAAACCAAGATTCATATAGTATGAAAAATGAACTTCATATTTCTTTTGGAAAAAGATGT 1323  
 |||||||  
 Db 2698 GAAACCAAGATTCATATAGTATGAAAAATGAACTTCATATTTCTTTTGGAAAAAGATGT 2757  
 |||||||  
 QY 1324 GGCATTTATTCAGTATTTTATATATAGACTTGGCTGACATTAATATAGTATTT 1383  
 |||||||  
 Db 2758 GGCATTTATTCAGTATTTTATATATAGACTTGGCTGACATTAATATAGTATTT 2817  
 |||||||  
 QY 1384 TGGACAAGAGTCTGCTGACAAAGCTATACCTAATATATAGTATATAAACATATAGATATG 1443  
 |||||||  
 Db 2818 TGGACAAGAGTCTGCTGACAAAGCTATACCTAATATATAGTATATAAACATATAGATATG 2877  
 |||||||  
 QY 1444 AGTCTTTGACAGTTTAACTCAATGAGATCAGAAATTTCTATGATTTGAGAAATGTTT 1503  
 |||||||  
 Db 2878 AGTCTTTGACAGTTTAACTCAATGAGATCAGAAATTTCTATGATTTGAGAAATGTTT 2937  
 |||||||  
 QY 1504 AATATCAATCTTAAATCTTGAATTTCTAAGAGCTTATTTGTTCTTTGGCTGATGTA 1563  
 |||||||  
 Db 2938 AATATCAATCTTAAATCTTGAATTTCTAAGAGCTTATTTGTTCTTTGGCTGATGTA 2997  
 |||||||  
 QY 1564 GTATATTTGAATTTGTTGAATTAATTAATTTCTCATTTGTAATAATATATATATG 1623  
 |||||||  
 Db 2998 GTATATTTGAATTTGTTGAATTAATTAATTTCTCATTTGTAATAATATATATATG 3057  
 |||||||  
 QY 1624 AATATATTTGATTTGATTAATCAATATAGATGTTGTTGATTTGTTGATTTGATTAATA 1683  
 |||||||  
 Db 3058 AATATATTTGATTTGATTAATCAATATAGATGTTGTTGATTTGTTGATTTGATTAATA 3117  
 |||||||  
 QY 1684 TCTGTTGATTTGATTTGATTAATTTGCTTAAATATTTGTAATCTTAATTTATTTTTC 1743  
 |||||||  
 Db 3118 TCTGTTGATTTGATTTGATTAATTTGCTTAAATATTTGTAATCTTAATTTATTTTTC 3177  
 |||||||  
 QY 1744 TCTGTTGATTTGATTTGATTAATTTGCTTAAATATTTTCTTCAAGAAATTTCTAT 1803  
 |||||||  
 Db 3178 TCTGTTGATTTGATTTGATTAATTTGCTTAAATATTTTCTTCAAGAAATTTCTAT 3237  
 |||||||  
 QY 1804 GCTATATTAATAATAATATTTACTG 1830  
 |||||||  
 Db 3238 GCTATATTAATAATAATATTTACTG 3264  
 |||||||

## RESULT 3

AAD24022  
 ID AAD24022 standard; cDNA; 1914 BP.

AC AAD24022;

XX 26-MAR-2002 (first entry)

DE Human protein phosphatase-4 cDNA.

KW Human; protein phosphatase; PP-4; immune system disorder; AIDS; allergy;

KW neurological disorder; developmental disorder; Alzheimer's disease;

KW cell proliferative disorder; Huntington's disease; arteriosclerosis;

KW renal tubular acidosis; gonadal dysgenesis; cancer; adenocarcinoma;

KW leukaemia; transgenic animal; gene therapy; ss.

OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FH CDS 203..1465  
 FT //tag= a  
 FT /product= "Protein phosphatase-4"  
 PN W0200196546-A2.  
 XX 20-DEC-2001.  
 PD 14-JUN-2001; 2001WO-US19442.  
 PF 16-JUN-2000; 2000US-212447P.  
 PR 22-JUN-2000; 2000US-213746P.  
 PR 29-JUN-2000; 2000US-215210P.  
 PR 06-JUL-2000; 2000US-216529P.  
 PR 12-JUL-2000; 2000US-218080P.  
 PR 21-JUL-2000; 2000US-220117P.  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX Au-Young J, Baughn MR, Ding L, Elliott VS, Gandhi AR, Griffin JA;  
 PI Hatalla A, Kearney L, Lee EA, Lu Y, Nguyen DB, Patterson C;  
 PI Rammumar J, Reddy R, Sanjanwala MS, Stewart EA, Tang YT;  
 PI Thornton M, Tribouley CM, Walla NK, Yang J, Yao MG, Yue H;  
 XX WPI: 2002-090206/12.  
 DR P-PSDB; AAE14454.  
 DR Novel polypeptide, useful for diagnosing, treating or preventing  
 PT disorders of growth and development, immune system, neurological and  
 PT cell proliferation diseases, comprises cancer protein phosphatase  
 PT polypeptides -  
 PS Claim 5; Page 112-113; 116pp; English.  
 XX The present sequence is human protein phosphatase (PP)-4 cDNA.  
 CC PP polynucleotide and polypeptide are useful in the diagnosis,  
 CC treatment and prevention of immune system disorders, neurological  
 CC disorders, developmental disorders and cell proliferative disorders.  
 CC Examples of immune system disorders include acquired immune deficiency  
 CC syndrome (AIDS), severe combined immunodeficiency disease (SCID),  
 CC adult respiratory distress syndrome, allergies, amyloidosis,  
 CC anaemia, asthma, arteriosclerosis, Crohn's disease, atopic  
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
 CC Graves' disease, multiple sclerosis, myasthenia gravis, myocardial or  
 CC pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,  
 CC psoriasis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome,  
 CC scleroderma, systemic sclerosis, trauma, neurological disorders include  
 CC Alzheimer's disease, Huntington's disease, dementia, epilepsy,  
 CC Parkinson's disease, mental retardation and other developmental  
 CC disorders of central nervous system such as Down's syndrome, cerebral  
 CC palsy, periodic paralysis, mental disorders including mood, anxiety,  
 CC and schizophrenic disorders, seasonal affective disorder such as  
 CC akathisia, amnesia, cataplexia, dyskinesia; developmental disorders  
 CC include e.g. renal tubular acidosis, Duchenne and Becker muscular  
 CC dystrophy, gonadal dysgenesis, hypothyroidism; cell proliferative  
 CC disorders include e.g. actinic keratosis, arteriosclerosis,  
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis and  
 CC cancer including adenocarcinoma, leukaemia. The polypeptide and  
 CC polynucleotide are further useful for analysing proteome of a tissue  
 CC or a cell type, for screening an agonist/antagonist, a compound that  
 CC specifically binds to it or its modulator. The polynucleotide is useful  
 CC for creating knockin humanised animals (pigs) or transgenic animals  
 CC (mice or rats) to model human disease, for generating a transcript image  
 CC of a tissue or cell type, which represents the global pattern of gene  
 CC expression by a particular tissue or cell type.  
 XX Sequence 1914 BP; 565 A; 378 C; 438 G; 533 T; 0 other;  
 SO Query Match 31.5%; Score 576.4; DB 24; Length 1914;  
 Best Local Similarity 99.8%; Pred. No. 5.2e-104;





Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.  
 OS Homo sapiens.  
 WO200060076-A2.  
 PD 12-OCT-2000.  
 PF 15-FEB-2000; 2000MO-US05308.  
 PR 02-APR-1999; 99US-0285480.  
 PR 23-JUN-1999; 99US-0339338.  
 PR 02-SEP-1999; 99US-0389681.  
 PR 03-NOV-1999; 99US-0433826.  
 (CORI-) CORIXA CORP.  
 PA Yugu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;  
 PI WPI; 2001-122627/13.  
 DR An isolated polypeptide useful for the treatment and diagnosis of  
 PT tumors e.g. breast cancer comprises at least an immunogenic portion of  
 PT a breast tumor protein.  
 PS Claim 6; Page 153; 238pp; English.  
 CC The present invention provides the coding sequences and some protein  
 CC sequences of proteins associated with breast cancer in humans. These  
 CC sequences can be used in the diagnosis and treatment of cancers,  
 CC particularly breast tumours.  
 XX Sequence 381 BP; 106 A; 68 C; 51 G; 142 T; 14 other;  
 SQ  
 Query Match 19.7%; Score 360.6; DB 22; Length 381;  
 Best Local Similarity 95.3%; Pred. No. 9.6e-62;  
 Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
 QY 949 AGAGGAGCTTGGCAGTCCGATGCTTCCATCCACCCCTTAACCTTAACATATTA 1008  
 DB 1 AGAGGAGATNTTGGCATGCTGATNTCTTCCATNCCACCTTAACATATTA 60  
 QY 1009 CTAGTGTCTTGTAAAGCAAGTATTAACCTTAACCTTACTCTTGGCCCTT 1068  
 DB 61 CTAGTGTCTTGTAAAGCAAGTATTAACCTTAACTTACTCTTGGCCCTT 120  
 QY 1069 TAGCTAATTAAGTGTGATATGAGCATTTATATATCTGAGTCATCATGTAT 1128  
 DB 121 TAGCTAATTAAGTGTGATATGAGCATTTATATATCTGAGTCATCATGTAT 180  
 QY 1129 CTCTCATGTTGATGATTTTCAACCTAAGATCTATGATAGTTTTCAGAGTTC 1188  
 DB 181 CTCTCATGTTGATGATTTTCAACCTAAGATCTATGATAGTTTTCAGAGTTC 240  
 QY 1189 ATTAATCATTTATTTCTTACTTCTGACCTCTGTCGTTGAACATTTAGAAATG 1248  
 DB 241 ATTAATCATTTATTTCTTACTTCTGACCTCTGTCGTTGAACATTTAGAAATG 300  
 QY 1249 GGAACCCCAATTTTGGAAACAGATTCATAGCAGTGAAGAACTTCATTTCTG 1308  
 DB 301 GGAACCCCAATTTTGGAAACAGATTCATAGCAGTGAAGAACTTCATTTCTG 360  
 QY 1309 TTTTGAAGAAGATGGCCAT 1329  
 DB 361 TTTTGAAGAAGATGGACCT 381  
 RESULT 6  
 AAV81744  
 ID AAV81744 standard; cDNA; 1785 BP.  
 AC AAV81744;  
 XX

DT 10-MAR-1999 (first entry)  
 DE Mouse PTP05 encoding cDNA.  
 XX  
 KW PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; ss.  
 XX  
 OS Mus sp.  
 FH Key Location/Qualifiers  
 FT CDS 199..1479  
 FT /tag= a  
 PN WO9849317-A2.  
 PD 05-NOV-1998.  
 PF 27-APR-1998; 98MO-US08439.  
 PR 23-OCT-1997; 97US-0063595.  
 PR 28-APR-1997; 97US-0044428.  
 PR 20-MAY-1997; 97US-0047222.  
 PR 11-JUN-1997; 97US-0049477.  
 PR 11-JUN-1997; 97US-0049756.  
 PR 18-JUN-1997; 97US-0049914.  
 (SUGEN-) SUGEN INC.  
 PA App H, Clary D, Courtneidge SA, Hui TH, Jallat B;  
 PI Markby D, Onrust S, Peles E, Plozman GD;  
 PI P-PDB: AAW89249.  
 DR WPI; 1999-009434/01.  
 DR P-PDB: AAW89249.  
 XX  
 PT New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease  
 PS Claim 2; Page 147; 193pp; English.  
 XX  
 CC The present invention describes isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC present sequence encodes mouse PTP05. The above proteins, other than  
 CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
 CC substances that modulate their activity (i.e. agonists and antagonists).  
 CC including NBP) in vivo or in vitro. These substances are used to treat  
 CC or prevent diseases associated with abnormal signal transduction  
 CC pathways that involve the proteins, particularly cancer (e.g. Leukemia  
 CC and Lymphoma), while modulators of ALK-7 (which is a type I receptor  
 CC serine/threonine kinase) are used to promote neuronal survival.  
 CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
 CC diseases. Nucleic acid fragments of the polynucleotides encoding the  
 CC proteins can be used as probes to identify and clone related sequences;  
 CC to detect protein-encoded RNA; to generate transgenic animals and in  
 CC gene therapy (optionally after mutation). Ab are used to determine the  
 CC proteins.  
 XX  
 SQ Sequence 1785 BP; 527 A; 361 C; 408 G; 489 T; 0 other;  
 Query Match 6.9%; Score 126.4; DB 20; Length 1785;  
 Best Local Similarity 64.6%; Pred. No. 1.2e-15;  
 Matches 301; Conservative 0; Mismatches 126; Indels 39; Gaps 6;  
 QY 102 CAGATACCGAGATATCTTCCATTTCAACATCAGATGATAGTGGCCCAAGAGAGAC 161  
 DB 1328 CTGCCATCGAGAGAGATCTTGTGACATTTATGACATGACCCAGATGAGAAAGC 1387  
 QY 162 AACGTTCTGGCATGTTTCAAGAGAGAGATGATCAGTTTGTACGATATTTGCTTG 221  
 DB 1388 AGCGCTGGCATGATTTCAACCAAGAGAGAGATGATCAGTTTGTATGAAATGTGCTTG 1447

QY 222 AAGTCTTCGAAACTTCTGACTTTGGATTAGAAAGACTTCTGCTCCTCACTTGA 281  
 DB 1448 AAGTCTTCGAAACTTCTGCTTTGTTATTAAAGAGACTTCTG-CGCTGCTCCTGAG 1506  
 QY 282 ATTACCAAGTGGTGTGACCTGCTATTAAGAAACATGTTGACAGTGTGCTGAAGGCTT 341  
 DB 1507 GTTACCGAGCAGCTTGGAGCC-----TGAGCCGCTGTGAGCG-TC 1546  
 QY 342 TGTATGCAATACATCTGCTTCTGTTATTCAGTTTATTTCTTCTTAAAGCTCCCT 401  
 DB 1547 TCGGGGCGCTGACGCTGCTTC-----TGATTTTCTCTCTGAAAGTCCC 1592  
 QY 402 GAAGGCAATATCATTTGCTTGGGCTGATCACTGTTACTTATGATCTTCTGAGCAA 461  
 DB 1593 TGAAGTACACTACTGCGGACAGAGTGAAGTCTTCCAC---TTGATCTTCTGAAACAA 1649  
 QY 462 TATCAAAATTAATCTCCACATTTTCCAGTGAACAGATTTACATAAAGATGACGT 521  
 DB 1650 GAGCAAAATTAACCT-CCATGCTCTTACGAAAGGAAAGTGTGATGAACACCTCCGCT 1708  
 QY 522 TGGCTATTTGTTGAAGGATTAACAGAGCCCAATTAAGATTAA 567  
 DB 1709 TGGCTGTCTGTTTGTGATTACAGAGCTTAATAAAGACTTGA 1754  
 RESULT 7  
 AAV81745  
 ID AAV81745 standard; cDNA; 1896 BP.  
 AC AAV81745;  
 DT 10-MAR-1999 (first entry)  
 DE Mouse PTP05 isoform #1 encoding cDNA.  
 KW PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; ss.  
 OS Mus sp.  
 FH Key location/Qualifiers  
 FT CDS 199..1590  
 FT /\*tag- a  
 XX MO9849317-A2.  
 PN 05-NOV-1998.  
 PD 27-APR-1998; 98MO-US08439.  
 PE 23-OCT-1997; 97US-0063595.  
 PR 28-APR-1997; 97US-0044428.  
 PR 20-MAR-1997; 97US-0047222.  
 PR 11-JUN-1997; 97US-0049477.  
 PR 11-JUN-1997; 97US-0049756.  
 PR 18-JUN-1997; 97US-0049914.  
 PA (SUGEN-) SUGEN INC.  
 PI App H, Clary D, Courtenidge SA, Hul TH, Jallal B;  
 PI Markby D, Onrust S, Peles E, Plozman GD;  
 DR WPI; 1999-009434/01.  
 DR P-PSDB; AAW89250.  
 XX New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease  
 PS Claim 2; Page 147-148; 193pp; English.  
 XX

CC The present invention describes proteins isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC present sequence encodes mouse PTP05. The above proteins, other than  
 CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
 CC substances that modulate their activity (i.e. agonists and antagonists,  
 CC including NBP) in vivo or in vitro. These substances are used to treat  
 CC or prevent diseases associated with abnormal signal transduction  
 CC pathways that involve the proteins, particularly cancer (e.g. leukaemia  
 CC and lymphoma), while modulators of ALK-7 (which is a type I receptor  
 CC serine/threonine kinase) are used to promote neuronal survival,  
 CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
 CC diseases. Nucleic acid fragments of the polynucleotides encoding the  
 CC proteins can be used as probes to identify and clone related sequences;  
 CC to detect protein-encoded RNA; to generate transgenic animals and in  
 CC gene therapy (optionally after mutation). Ab are used to determine the  
 CC proteins.  
 SQ Sequence 1896 BP; 572 A; 372 C; 427 G; 525 T; 0 other;  
 Query Match 6.9%; Score 126.4; DB 20; Length 1896;  
 Best Local Similarity 64.6%; Pred. No. 1.2e-15;  
 Matches 301; Conservative 0; Mismatches 126; Indels 39; Gaps 6;  
 QY 102 CAGATACCGAGATATCTTCATTTCAATCATCATGATGATAGTGCCCAATGAGAAAC 161  
 DB 1439 CTGCCATCGAGAGAGAACTCTCTTTGACATTAATGAACATGAGACCATGAGAAAC 1498  
 QY 162 AAGCTTGGCATGCTTCAACAGAGAGAGATATCATTTGTTACGATATGCTTGG 221  
 DB 1499 AGCGCTGGGCAATGATTCAAACAGAGAGAGATGATGATGATGATGATGATGATG 1558  
 QY 222 AAGTCTTCGAAACTTCTGACTTTGGATTAGAAAGACTTCTGCTCCTCACTTGA 281  
 DB 1559 AAGTCTTCGAAACTTCTGCTTTGTTATTAAAGAGACTTCTG-CGCTGCTCCTGAG 1617  
 QY 282 ATTACCAAGTGGTGTGACCTCCTCATTAAGAAACATGTTGACGTGTGCTGAAGGCTT 341  
 DB 1618 GTTACCGAGCAGCTTGGAGCC-----TGAGCCGCTGTGAGCG-TC 1657  
 QY 342 TGTATGCAATACATCTGCTTCTTGTGTTATCATGTTTATTTCTTCTTAAAGCTCCCT 401  
 DB 1658 TCGGGGCGCTGACGCTGCTTC-----TGATTTTCTCTCTGAAAGTCCC 1703  
 QY 402 GAAGGCAATATCATTTGCTTGGGCTGATCACTGTTACTTATGATCTTCTGAGCAA 461  
 DB 1704 TGAAGTACACTACTGCGGACAGAGATGAACCTTTCCAC---TTGATCTTCTGAAACAA 1760  
 QY 462 TATCAAAATTAATCTCCACATTTTCCAGTGAACAGATTTACATAAAGATGACGT 521  
 DB 1761 GAGCAAAATTAACCT-CCATGCTCTTACGAAAGGAAAGTGTGATGAACACCTCCGCT 1819  
 QY 522 TGGCTATTTGTTGAAGGATTAACAGAGCCCAATTAAGATTAA 567  
 DB 1820 TGGCTGTCTGTTTGTGATTACAGAGCTTAATAAAGACTTGA 1865  
 RESULT 8  
 ABL55644/C  
 ID ABL55644 standard; DNA; 50000 BP.  
 AC ABL55644;  
 DT 01-JUL-2002 (first entry)  
 DE AMEPV genome fragment#2.  
 DE AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;  
 KW genetic deficiency disorder; ds.  
 OS Amsacta moorei entomopoxvirus.  
 PS WO200212526-A2.  
 XX



of the printed specification, but was obtained in electronic format directly from WIPO at ftp://wipo.int/pub/published-pct\_sequences.

Sequence 5413 BP; 1391 A; 65 C; 1102 G; 2855 T; 0 other;

Query Match 4.3%; Score 79.4; DB 22; Length 5413;  
Best Local Similarity 47.7%; Pred. No. 2.4e-06;  
Matches 359; Conservative 0; Mismatches 386; Indels 8; Gaps 4;

```

QY 1080 TAAAGTTGATATAGGCAATTAATTAATCTGAGTCATTCAGTGTCTCATGTTT 1139
DB 1233 TAAAGTTGATATAGGCAATTAATTAATCTGAGTGTCTCATGTTT 1139
QY 1140 GATGATATTTTCAACATAGATGATGATGATGATGATGATGATGATGATGAT 1199
DB 1293 GATGATATTTTCAACATAGATGATGATGATGATGATGATGATGATGATGAT 1199
QY 1200 TATTTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1259
DB 1353 TATATATATATATATATATATATATATATATATATATATATATATATATAT 1411
QY 1260 TTGGGAAACGAGATTCATGATGATGATGATGATGATGATGATGATGATGAT 1319
DB 1412 TTTATATAGGATATATATATATATATATATATATATATATATATATATAT 1471
QY 1320 AGTGGCCATTTATAGATATTTTATATATAGATGATGATGATGATGATGATG 1379
DB 1472 GATATATATATATATATATATATATATATATATATATATATATATATATAT 1531
QY 1380 ATTTTGACACACACACACACACACACACACACACACACACACACACACACAC 1434
DB 1532 ATTTATATATATATATATATATATATATATATATATATATATATATATATAT 1591
QY 1435 ATATATATATATATATATATATATATATATATATATATATATATATATAT 1494
DB 1592 TTTATATATATATATATATATATATATATATATATATATATATATATATAT 1651
QY 1495 AAAATGTTTATATATATATATATATATATATATATATATATATATATATAT 1553
DB 1652 TATATATATATATATATATATATATATATATATATATATATATATATATAT 1711
QY 1554 GGCTGATATATATATATATATATATATATATATATATATATATATATATAT 1613
DB 1712 ATTTATATATATATATATATATATATATATATATATATATATATATATATAT 1771
QY 1614 ATATGCAAAATATATATATATATATATATATATATATATATATATATATAT 1673
DB 1772 ATTTTATATATATATATATATATATATATATATATATATATATATATATAT 1831
QY 1674 TCA-ATATATATATATATATATATATATATATATATATATATATATATAT 1732
DB 1832 TTTATATATATATATATATATATATATATATATATATATATATATATATAT 1891
QY 1733 ATTTTATATATATATATATATATATATATATATATATATATATATATATAT 1792
DB 1892 TATTTTATATATATATATATATATATATATATATATATATATATATATATAT 1951
QY 1793 AAGATATATATATATATATATATATATATATATATATATATATATATAT 1855
DB 1952 AGTATATATATATATATATATATATATATATATATATATATATATATAT 1984

```

RESULT 10  
AAV81746  
ID AAV81746 standard; cDNA; 1692 BP.

AC AAV81746;  
XX 10-MAR-1999 (first entry)  
DT Mouse PTP05 isoform #2 encoding cDNA.  
XX

PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase; type I receptor serine/threonine kinase; cancer; leukemia; lymphoma; neurodegenerative disease; neuronal survival; Alzheimer's disease; Parkinson's disease; Huntington's disease; ss.

Mus sp.

Key Location/Qualifiers  
FT 199..1413  
FT CDS /tag=a  
FT /note="no stop codon given"

MO9849317-A2.

05-NOV-1998.

27-APR-1998; 98MO-US08439.

23-OCT-1997; 97US-0063595.

28-APR-1997; 97US-0044428.

20-MAY-1997; 97US-0047222.

11-JUN-1997; 97US-0048477.

11-JUN-1997; 97US-0048756.

18-JUN-1997; 97US-0049914.

(SUGEN) SUGEN INC.

App H, Clary D, Courtneidge SA, Hui TH, Jallat B;

Markby D, Onrust S, Peles E, Plozman GD;

WPI; 1999-009434/01.

P-PSDB; AAM89251.

New nucleic acid encoding specific protein tyrosine phosphatases -

PT useful for identifying specific modulators for treatment and

prevention of cancer and neurodegenerative disease

Claim 2; Page 148-149; 193pp; English.

The present invention describes isolated, enriched or purified nucleic acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The present sequence encodes mouse PTP05. The above proteins, other than ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify substances that modulate their activity (i.e. agonists and antagonists, including NBP) in vivo or in vitro. These substances are used to treat or prevent diseases associated with abnormal signal transduction pathways that involve the proteins, particularly cancer (e.g. leukemia and lymphoma), while modulators of ALK-7 (which is a type I receptor serine/threonine kinase) are used to promote neuronal survival, particularly for treating Alzheimer's, Parkinson's or Huntington's diseases. Nucleic acid fragments of the polynucleotides encoding the proteins can be used as probes to identify and clone related sequences; CC to detect protein-encoded RNA; to generate transgenic animals and in gene therapy (optionally after mutation). Ab are used to determine the proteins.

Sequence 1692 BP; 507 A; 342 C; 386 G; 457 T; 0 other;

Query Match 4.3%; Score 78.6; DB 20; Length 1692;  
Best Local Similarity 79.5%; Pred. No. 2.9e-06;  
Matches 93; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

QY 12 TGATATCATGCGAATTTATGCTTTAGAACTTAAGAAATGCTGATGATCAATC 71
DB 687 TGATATCATGCGTGAATTTTGGAGTTAGAACTTCTGATGATCAATTC 746
QY 72 TGGGATCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 128
DB 747 TGGGATCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 803

```

RESULT 11  
AAS46758

ID AAS46758 standard; DNA; 5447 BP.  
 AC AAS46758;  
 XX  
 DT 18-DEC-2001 (first entry)  
 DE Tumour suppressor gene derived chemically modified sequence #482.  
 DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
 KW cancer; Cpg dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 OS Homo sapiens.  
 XX  
 XX WO200168912-A2.  
 PD 20-SEP-2001.  
 PF 15-MAR-2001; 2001WO-EP02955.  
 PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Plepenbrock C, Berlin K;  
 XX  
 XX WPI; 2001-602752/68.  
 PT Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer  
 XX  
 XX Claim 1; SEQ ID No 482; 27pp; English.  
 XX  
 XX The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
 CC peptide nucleic acid oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with Cpg dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific  
 CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes. Sequences with even numbered Seq ID numbers are the  
 CC complementary sequence of the corresponding odd numbered sequence (e.g.  
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
 CC is missing).  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from Wipo at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 5447 BP; 1427 A; 183 C; 1297 G; 2540 T; 0 other;  
 SQ  
 Query Match 4.3%; Score 78; DB 22; Length 5447;  
 Best Local Similarity 47.8%; Pred. No. 4.5e-06;  
 Matches 260; Conservative 0; Mismatches 280; Indels 4; Gaps 1;  
 OY 1285 GAAATGAAATCTTCATATTTGTTTGAAGAAAGATGCGCCATTATACAGTAATTTT 1344

DB 2394 GTATTGAAAGAAATGTTAAATTTTGAATTCGAATTTTGTATTAATAAGCAATTTT 2453  
 OY 1345 ATTATGACCTTGGCTGCTACATTAATAGATATTT---TGACACAGAGATCTCG 1400  
 DB 2454 TTAATAAGATGAGAGATTTTAATGATTAATCGTATGTAAGGGAATTTGTAAAA 2513  
 OY 1401 TGACAGCTATACCTAATTAATTAAGCTATTAACAAATAGATAGAGTGTGTACAGTTA 1460  
 DB 2514 AAAAAAAAAAAAAAAAAAGTTAATTAATGAGTAAATTAATTTGTAAGTTGT 2573  
 OY 1461 ACTCAATGAGATCGCAATATCTATGTATGAGAAATGTTAATATCAATCTATAAAT 1520  
 DB 2574 TTAATATTGAAATGATTAATTAATTTGTTTATTAATAAAGTTAAGCTATTGTAAGT 2633  
 OY 1521 CTTGAATTTCTAAGAGGCTTATTTTGTCTTGGCTGATGAGTAAATTTGATTTG 1580  
 DB 2634 TTTAATTTTATTAAGTTTAAAAAGTTAATTTTAAAAAGAGTTAAATAAATA 2693  
 OY 1581 GAATAATTAATTAATTTCTCATTTGTAATAATTAATGCAAAATATATTGATGTTA 1640  
 DB 2694 TTGATTTGAATAATTAATTAATTTATGTTGTAATAAATTTATTTATTTAGATAT 2753  
 OY 1641 ATCAATATAGATCTGCTTTACATGTTCAATGATATTAATTCGTAAATTCATT 1700  
 DB 2754 TATTAATTAAGGATTTGATTAATTAATGATGTTTGAATTTATGTTTATTAAT 2813  
 OY 1701 TTGATTAATGCGCTTAAATTTGTAATCTAAATTTATTTCTGCTGTTACTGTAAA 1760  
 DB 2814 AATTAATTAATTTATTTTATTTTATTTTATTTTATTTTATTAATTTTATTTTGTG 2873  
 OY 1761 TATACCTATATGATTAACAAATTTCTGAGAAATTCATGCTATTAATAATAA 1820  
 DB 2874 TTAATGTAATAATTTATTAATTAATTTTATTAATTTTATTAATAAGACAAA 2933  
 OY 1821 ATAT 1824  
 DB 2934 ATAT 2937

RESULT 12  
 ABL92318  
 ID ABL92318 standard; DNA; 7334 BP.  
 XX  
 AC ABL92318;  
 XX  
 DT 01-JUL-2002 (first entry)  
 DE Chemically treated DNA repair gene fragment#64.  
 DE DNA repair; cytosine methylation; PMS2L1; PMS2L2; PMS2L3;  
 KW PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; PDG; INPBL1; RFC4;  
 KW DITTL; FANCB; XRC8; ataxia telangiectasia; aging; Bloom's syndrome;  
 KW Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;  
 KW immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;  
 KW cancer; ds.  
 OS unidentified.  
 OS  
 OS WO200181622-A2.  
 PN  
 XX  
 PD 01-NOV-2001.  
 PF 06-APR-2001; 2001WO-EP03972.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Plepenbrock C, Berlin K;









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 07:34:22 ; Search time 4683 Seconds

(without alignments)  
11365.082 Million cell updates/sec

Title: US-09-935-703-2

Perfect score: 1832

Sequence: 1 ggcacgaggtatcatatcat.....aaaataaatactactgtg 1832

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb\_da:\*

2: gb\_hg:\*

3: gb\_in:\*

4: gb\_ov:\*

5: gb\_ov:\*

6: gb\_ov:\*

7: gb\_ov:\*

8: gb\_ov:\*

9: gb\_ov:\*

10: gb\_ov:\*

11: gb\_ov:\*

12: gb\_ov:\*

13: gb\_ov:\*

14: gb\_ov:\*

15: gb\_ov:\*

16: gb\_ov:\*

17: gb\_ov:\*

18: gb\_ov:\*

19: gb\_ov:\*

20: gb\_ov:\*

21: gb\_ov:\*

22: gb\_ov:\*

23: gb\_ov:\*

24: gb\_ov:\*

25: gb\_ov:\*

26: gb\_ov:\*

27: gb\_ov:\*

28: gb\_ov:\*

29: gb\_ov:\*

30: gb\_ov:\*

31: gb\_ov:\*

32: gb\_ov:\*

33: gb\_ov:\*

34: gb\_ov:\*

35: gb\_ov:\*

36: gb\_ov:\*

37: gb\_ov:\*

38: gb\_ov:\*

39: gb\_ov:\*

40: gb\_ov:\*

41: gb\_ov:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1822.4	99.5	2499	9	BC036539	BC036539 Homo sapi
2	1688.6	92.2	2786	9	HSMB000374	AL050040 Homo sapi
3	1645.4	88.8	197389	9	AL358791	AL358791 Human DNA
4	1642.2	89.6	65229	9	AL672108	AL672108 Human DNA
5	576.4	31.5	1914	6	AX354535	AX354535 Sequence
6	360.6	19.7	381	6	AX302888	AX302888 Sequence
7	308	16.8	3090	10	D64141	D64141 Mus musculu
8	170.8	9.3	173224	2	AC130106	AC130106 Rattus no
9	106.6	5.8	55785	2	AC118715	AC118715 Mus muscu
10	97.6	5.3	137889	9	AC073269	AC073269 Homo sapi
11	95	5.2	103737	9	AL603966	AL603966 Human DNA
12	95	5.2	171930	9	AL450334	AL450334 Human DNA
13	86.4	4.7	111861	9	AC069435	AC069435 Homo sapi
14	85.2	4.7	104992	2	AC005504	AC005504 Plasmodiu
15	85.2	4.7	169546	2	AC004157	AC004157 Plasmodiu
16	84.8	4.6	86827	3	PFMAL3P5	AL034556 Plasmodiu
17	84.6	4.6	178785	9	AC068139	AC068139 Homo sapi
18	84.4	4.6	15421	3	PFCOMP1RA	X95275 P. falciparu
19	83.2	4.5	50000	6	AX392734	AX392734 Sequence
20	83.2	4.5	232392	14	AF250284	AF250284 Amsacta m
21	81.6	4.5	162445	9	AL158151	AL158151 Human DNA
22	81.6	4.4	178273	2	AC005308	AC005308 Plasmodiu
23	80.8	4.4	148075	2	AC110014	AC110014 Homo sapi
24	80.8	4.4	177293	2	AC069525	AC069525 Homo sapi
25	80.6	4.4	17443	3	AC024806	AC024806 Caenorhab
26	80.6	4.4	116696	3	PFMAL3P3	298547 Plasmodiu
27	80.4	4.4	72243	9	AL731858	AL731858 Human DNA
28	80.4	4.4	172816	9	AC093899	AC093899 Homo sapi
29	80.2	4.4	127820	9	HSJ612B15	AL049597 Human DNA
30	79.8	4.4	143802	9	AL161431	AL161431 Human DNA
31	79.8	4.4	176174	9	AC007483	AC007483 Homo sapi
32	79.6	4.3	137411	9	AC092066	AC092066 Homo sapi
33	79.4	4.3	5413	6	AX251450	AX251450 Sequence
34	79	4.3	155106	9	AC104069	AC104069 Homo sapi
35	79	4.3	161699	9	AC107300	AC107300 Homo sapi
36	78.8	4.3	185596	9	AC021553	AC021553 Homo sapi
37	78.2	4.3	328866	2	AC107203	AC107203 Homo sapi
38	78	4.3	5447	6	AX251516	AX251516 Sequence
39	77.8	4.2	1192	9	HSJ323759	AJ323759 Homo sapi
40	77.8	4.2	96953	9	AC079621	AC079621 Homo sapi
41	77.6	4.2	6591	8	YSCWTCG06	L36890 Saccharomyc
42	77.6	4.2	15635	3	AB083339	AB083339 Bombyx mo
43	77.6	4.2	15656	3	AB070264	AB070264 Bombyx mo
44	77.6	4.2	152209	9	HS1108011	AL034419 Human DNA
45	77.6	4.2	169422	9	AC093309	AC093309 Homo sapi

#### ALIGNMENTS

RESULT 1  
BC036539  
LOCUS Homo sapiens, similar to protein-tyrosine-phosphatase homolog  
DEFINITION DKFZP566K0524.1 - human (fragment), clone MGC:33863 IMAGE:5262458,  
BC036539  
ACCESSION BC036539  
VERSION BC036539.1 GI:22328116  
KEYWORDS mRNA, complete cds.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2499)  
AUTHORS Strausberg, R.

Pred. No. is the number of results predicted by chance to have a

**TITLE** Direct Submission  
**JOURNAL** Submitted (16-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK** NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shizaki Toshuyuki and Piero Carninci (RIKEN)  
 DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 contact: amandaregistry@systemsbiology.org  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

**FEATURES**  
 source  
 1. 2499  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MGC:33863 IMAGE:5262458"  
 /tissue="type="Brain, hippocampus"  
 /clone\_id="NIH\_MGC\_95"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"  
 192..665  
 /codon\_start=1  
 /product="similar to protein-tyrosine-phosphatase homolog dKpfp566K0524.1 - human (fragment)"  
 /protein\_id="AAH36539.1"  
 /db\_xref="GI:22328117"  
 /db\_xref="LOCUSID:257430"  
 /translation="MSPRDRAEPVNDYEGDSAEADINFEITLPSSQENTPSPKVFENYNSKVLRLNPEHNDYEDVEEPESSGDSMTAGPERRDWSSEDEAA GFSOLSLISGTRKIVSEGLDQIAQIRPLIFNHEQTAIDCLKILEKNSV"

**CDS**  
 766 a 438 c 486 g 809 t

**BASE COUNT**  
 766 a 438 c 486 g 809 t

**ORIGIN**  
 Query Match 99.5%; Score 1822.4; DB 9; Length 2499;  
 Best Local Similarity 99.9%; Pred. No. 4.4e-283;  
 Matches 1823; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 960 TAAAGACATGTTGACATGCTGCTGAAGGGCTTGTGATGACATCAATCTGCTTTGG 1019  
 QY TTTATCAGTTATTTCTTTCTTAAAGCTCCGAAGGCAATATCTTGGCTGGGT 428  
 Db 1020 TTTATCAGTTATTTCTTTCTTAAAGCTCCGAAGGCAATATCTTGGCTGGGT 1079  
 QY GATCAGCTGTTTACTTATGATCTTGTGAGCAATATCAATAATCTCCACATTTTCA 488  
 Db 1080 GATCAGCTGTTTACTTATGATCTTGTGAGCAATATCAATAATCTCCACATTTTCA 1139  
 QY GTGAACACATGTTTACTTATGATCTTGTGAGCAATATCAATAATCTCCACATTTTCA 548  
 Db 1140 GTGAACACATGTTTACTTATGATCTTGTGAGCAATATCAATAATCTCCACATTTTCA 1199  
 QY GCCCATTAAGATTTAAATATATTCATTAAGATTTTATTTGAAAGTGGCTGAGAG 608  
 Db 1200 GCCCATTAAGATTTAAATATATTCATTAAGATTTTATTTGAAAGTGGCTGAGAG 1259  
 QY AGCTGAGATTTTCCAGGACTTTGTAACTTATCTGAGAGAAACATTAAGCCATATC 668  
 Db 1260 AGCTGAGATTTTCCAGGACTTTGTAACTTATCTGAGAGAAACATTAAGCCATATC 1319  
 QY ATGACCTCTTCCAGGACTTTTAAAGACAGATGCTATCTATCTTCTTACTAGAGCTG 728  
 Db 1320 ATGACCTCTTCCAGGACTTTTAAAGACAGATGCTATCTATCTTCTTACTAGAGCTG 1379  
 QY TACTTTTGTGCTGATTTGAATTAACCATTTTAAAGAGTCCAGTTAGGCTGATTAAC 788  
 Db 1380 TACTTTTGTGCTGATTTGAATTAACCATTTTAAAGAGTCCAGTTAGGCTGATTAAC 1439  
 QY TTTGACACAAATTTGCTTCCATTTCTTCAATTTTCAATTTCTGCTTCTTCAAGCTGC 848  
 Db 1440 TTTGACACAAATTTGCTTCCATTTCTTCAATTTTCAATTTCTGCTTCTTCAAGCTGC 1499  
 QY TTAGACCAAGCCTGTGCTGAGGACATTTTCAATTTGAGGAGAGCTCTCCGAGTGA 908  
 Db 1500 TTAGACCAAGCCTGTGCTGAGGACATTTTCAATTTGAGGAGAGCTCTCCGAGTGA 1559  
 QY ACAGTCCAAATTAATATAGATTTTATATAGAAAGCCAAAGAGAGCTTTGCCATGC 968  
 Db 1560 ACAGTCCAAATTAATATAGATTTTATATAGAAAGCCAAAGAGAGCTTTGCCATGC 1619  
 QY CTGAGTCTTCTTCTTCCATTTCCACCTTCAACCTTCAATATTTACTTGTCTTTTAAA 1679  
 Db 1620 CTGAGTCTTCTTCTTCCATTTCCACCTTCAACCTTCAATATTTACTTGTCTTTTAAA 1679  
 QY GCAAGTATTCCTTAACTTCCCTTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1088  
 Db 1680 GCAAGTATTCCTTAACTTCCCTTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1739  
 QY ATATAGGATTTATATATATTTGATGATCTTCAATTTGATGATGATGATGATGATGAT 1148  
 Db 1740 ATATAGGATTTATATATATTTGATGATCTTCAATTTGATGATGATGATGATGATGAT 1799  
 QY TTTCAAACTAAGATCTATGATGATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1208  
 Db 1800 TTTCAAACTAAGATCTATGATGATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1859  
 QY TACTTCTCAGCTCTGTTGAACATTTAGAACTGATTTGGAACCAATTTTGGAAAA 1268  
 Db 1860 TACTTCTCAGCTCTGTTGAACATTTAGAACTGATTTGGAACCAATTTTGGAAAA 1919  
 QY CCAGATTCATAGTCATGAAATGGAACCTTCCATATTTCTGTTTGAAGATGTGGCA 1328  
 Db 1920 CCAGATTCATAGTCATGAAATGGAACCTTCCATATTTCTGTTTGAAGATGTGGCA 1979  
 QY TTTATCAGTAAATTTTATATAGGATTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1388  
 Db 1980 TTTATCAGTAAATTTTATATAGGATTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2039  
 QY AAGGAGTTCTGAGCAAGCTATACCTAATTAATTAAGCTATTAAGCTATTAAGCTAT 1448  
 Db 2040 AAGGAGTTCTGAGCAAGCTATACCTAATTAATTAAGCTATTAAGCTATTAAGCTAT 2099

OY 1449 TTGACAGTTTAACTCAATGAGATCAGATATATCTATGATTTGAGAAAAGTTTAAAT 1508  
 |||||||  
 DB 2100 TTGTACAGTTTAACTCAATGAGATCAGATATATCTATGATTTGAGAAAAGTTTAAAT 2159  
 |||||||  
 OY 1509 CAATCTAATAATCTGAAATTTCTAGAGAGCTTATTTGCTTTTGGCTGAAGAGATATA 1568  
 |||||||  
 DB 2160 CAATCTAATAATCTGAAATTTCTAGAGAGCTTATTTGCTTTTGGCTGAAGAGATATA 2219  
 |||||||  
 OY 1569 TTGCAATTTGCTGAATTAATTAATATCTCATTTGTAATAAATAATTAATATGCCAAAATAT 1628  
 |||||||  
 DB 2220 TTGCAATTTGCTGAATTAATTAATATCTCATTTGTAATAAATAATTAATATGCCAAAATAT 2279  
 |||||||  
 OY 1629 ATTGATGTTAATCAATAGATGATGCTGTTACATTTGCTCATATGAAATAATATGCT 1688  
 |||||||  
 DB 2280 ATTGATGTTAATCAATAGATGATGCTGTTACATTTGCTCATATGAAATAATATGCT 2339  
 |||||||  
 OY 1689 GTTAATTTCAATTTGTAATATTTGCTTAAATTTGTAATTTGTAATTTTCTCTCT 1748  
 |||||||  
 DB 2340 GTTAATTTCAATTTGTAATATTTGCTTAAATTTGTAATTTTCTCTCTCT 2399  
 |||||||  
 OY 1749 GTTACTGTAATAATAGTATATATATATATATATTTCTCAGAGAAATTTCTAT 1808  
 |||||||  
 DB 2400 GTTACTGTAATAATAGTATATATATATATATATTTCTCAGAGAAATTTCTAT 2459  
 |||||||  
 OY 1809 TATTAAATAATAATTTTACTGTC 1832  
 |||||||  
 DB 2460 TATTAAATAATAATTTTACTGTC 2483  
 |||||||

RESULT 2  
 HSM800374  
 LOCUS HSM800374 2786 bp mRNA linear PRI 18-FEB-2000  
 DEFINITION Homo sapiens mRNA; cDNA DKFZP566K0524 (from clone DKFZP566K0524);  
 partial cds.  
 ACCESSION AL050040  
 VERSION AL050040.1 GI:4884281  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 2786)  
 AUTHORS Angorge, W., Winkner, U., Mewes, H. W., Gassenhuber, J. and Wiemann, S.  
 TITLE Direct Subcloning  
 JOURNAL Submitted (15-MAY-1999) MIPS, Am Klopferstr. 18a, D-82152  
 Martinsried, GERMANY  
 COMMENT  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
 Heidelberg/Germany) within the cDNA sequencing consortium of the  
 German Genome Project.  
 This clone (DKFZP566K0524) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further  
 information about the clone and the sequencing project is available  
 at <http://www.mips.biochem.mpg.de/proj/cDNA/>.  
 Location/Qualifiers  
 source  
 1. 2786  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKFZP566K0524"  
 /tissue\_type="kidney"  
 /clone\_id="566 (synonym: hfkx2). Vector pAMP1; host  
 XL-2Blue; sites NotI + SalI"  
 /dev\_stage="fetal"  
 1. 1198  
 /gene="DKFZP566K0524"  
 1. 1198  
 /gene="DKFZP566K0524"  
 /note="strong similarity to protein-tyrosine-phosphatases"  
 /codon\_start=2  
 /product="hypothetical protein"

/protein\_id="CAB43248.1"  
 /db\_xref="GI:4884282"  
 /db\_xref="SPTREMBL:O9Y406"  
 /translation="GPNRRGCGGCTGGAAPQAVFENKYNSEKVLIRFPHNDY  
 EDVEESESSESDPSMTWARGPFRDRNRSSSEDEEAGSQAISPLSDTRKIVSEBEL  
 DOLAQIRPLIFNEHOITAIKDCLKILEKTAAYDIDMOEFMALELNLPDEFSGNOPS  
 NREKNRYDILPYDSTRVPLKSKDYINASTYRIVNCEGEYFIATOGPLSDIFW  
 OMVLENNSTVAMITREMGGIIKCYHPIKRPLEKHFRELVENLOILOYFIR  
 MFOYVEKSTGSHSYKQLOFTKMGPHGHPASDSFIRKIRARKSHLNGPMVHCASG  
 IGRGVFELCVIVFCAIVKDCSFSNIMDIYVQNRBSGNVQKREYHHCYDILVELVR  
 KLTLVD"  
 polyA\_signal  
 2761..2766  
 polyA\_site  
 2776  
 BASE COUNT 841 a 494 c 550 g 901 t  
 ORIGIN  
 Query Match 92.2%; Score 1688.6; DB 9; Length 2786;  
 Best Local Similarity 99.7%; Pred. No. 1.2e-261;  
 Matches 1702; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

OY 125 TTCAACATCATGATATAGTGGCCCAATAGAGAAACAACCTTCGCGATGGTTCAAGC 184  
 |||||||  
 DB 1070 TTCAACATCATGATATAGTGGCCCAATAGAGAAACAACCTTCGCGATGGTTCAAGC 1129  
 |||||||  
 OY 185 AAGGACAGTATCACTTTTGTACATATTTGCTGAAGTCTTCGAAACTTCGACT 244  
 |||||||  
 DB 1130 AAGGACAGTATCACTTTTGTACATATTTGCTGAAGTCTTCGAAACTTCGACT 1189  
 |||||||  
 OY 245 TTGATTAAGAAAGACTCTGTGCTCTCACTTGAATTAACAGTGGTTGACCTC 304  
 |||||||  
 DB 1190 TTGATTAAGAAAGACTCTGTGCTCTCACTTGAATTAACAGTGGTTGACCTC 1249  
 |||||||  
 OY 305 CTCATAAAGAACATTTTGCACCTGCTGAAGGCTTTCCTATGATTAATCTGCTTC 364  
 |||||||  
 DB 1250 CTCATAAAGAACATTTTGCACCTGCTGAAGGCTTTCCTATGATTAATCTGCTTC 1309  
 |||||||  
 OY 365 TTGGTTATCACTTTATTTCTTTCTAAAGCTCCCTGAAGGCAATATCATTTGGCTTC 424  
 |||||||  
 DB 1310 TTGGTTATCACTTTATTTCTTTCTAAAGCTCCCTGAAGGCAATATCATTTGGCTTC 1369  
 |||||||  
 OY 425 GGGTATCATGTTTACTTATGATCTGCTGAGCAATATCAATTAATCTCCACATTT 484  
 |||||||  
 DB 1370 GGGTATCATGTTTACTTATGATCTGCTGAGCAATATCAATTAATCTCCACATTT 1429  
 |||||||  
 OY 485 TCCAGTGAACAGATGTATCAATTAACAGATTCAGCTGGCTGATTTGGTGAAGGATTA 544  
 |||||||  
 DB 1430 TCCAGTGAACAGATGTATCAATTAACAGATTCAGCTGGCTGATTTGGTGAAGGATTA 1489  
 |||||||  
 OY 545 CAGAGCCCAATTAAGATTAATATATATATATATATATTTTGAAGAGTGGCTGC 604  
 |||||||  
 DB 1490 CAGAGCCCAATTAAGATTAATATATATATATATATATTTTGAAGAGTGGCTGC 1549  
 |||||||  
 OY 605 AGAGAGCTGAGGATTTCCAGACTTTGTAAGTCTTATTTCTGGAGACATAGGCCAAT 664  
 |||||||  
 DB 1550 AGAGAGCTGAGGATTTCCAGACTTTGTAAGTCTTATTTCTGGAGACATAGGCCAAT 1609  
 |||||||  
 OY 665 AATCATGACCTTTCAGAGCATTTTGAAGACAGATGTATCAATCTTCTTGAAGTGA 724  
 |||||||  
 DB 1610 AATCATGACCTTTCAGAGCATTTTGAAGACAGATGTATCAATCTTCTTGAAGTGA 1669  
 |||||||  
 OY 725 CCTGATCTTTTGTGCTGATTTGAATTAACAGTGAAGAGTGGCTGAGC 784  
 |||||||  
 DB 1670 CCTGATCTTTTGTGCTGATTTGAATTAACAGTGAAGAGTGGCTGAGC 1729  
 |||||||  
 OY 785 TAACCTTGAGACAAATTTGGCTTCATTTCTACATTTTCAATCTGCTCTCTACAG 844  
 |||||||  
 DB 1730 TAACCTTGAGACAAATTTGGCTTCATTTCTACATTTTCAATCTGCTCTCTACAG 1789  
 |||||||  
 OY 845 CTGCTAGACCAAGACCTGTGTCTGGAGACATTTCAATGATAGGAGAGAGCTCTCTG 904  
 |||||||  
 DB 1790 CTGCTAGACCAAGACCTGTGTCTGGAGACATTTCAATGATAGGAGAGAGCTCTCTG 1849  
 |||||||  
 OY 905 GTGAACAGTCCAAACATAATATGATGTTATATGAAGGCCAAGAGAGACATTTTGGC 964  
 |||||||





DB 144993 CAGAGGAAACAGATGTATACATAAAGCATGCGAGCTTGCTGATTTGGTTGAGGAGATAC 144934  
 QY 546 AGAGCCCAATTAAGATTTAAATATATCATTAAGATTTATTTGAAAAGGTGGCTGGA 605  
 DB 144933 AGAGCCCAATTAAGATTTAAATATATCATTAAGATTTATTTGAAAAGGTGGCTGGA 144874  
 QY 606 GAGAGCTGAGATTTCCAGAGCTTTGTAAGTCTTATTCGGGAGACATTAAGCCATA 665  
 DB 144873 GAGAGCTGAGATTTCCAGAGCTTTGTAAGTCTTATTCGGGAGACATTAAGCCATA 144814  
 QY 666 ATCATGACCTCTCCAGCATTTTAAAGACAGATGCTATTCATGTTCTTAGCTAGAGC 725  
 DB 144813 ATCATGACCTCTCCAGCATTTTAAAGACAGATGCTATTCATGTTCTTAGCTAGAGC 144754  
 QY 726 CTGTAATTTTGTGCTGCAATTTGAATAACCAAGTTTAAAAAGAGTCCAGTTAAGGTGACT 785  
 DB 144753 CTGTAATTTTGTGCTGCAATTTGAATAACCAAGTTTAAAAAGAGTCCAGTTAAGGTGACT 144694  
 QY 786 AACTTTGGACAAATTTGGCTTCATTTCTACATTTTCTAGTCTGCTCTCTACAGC 845  
 DB 144693 AACTTTGGACAAATTTGGCTTCATTTCTACATTTTCTAGTCTGCTCTCTCTACAGC 144634  
 QY 846 TGTAGACCAAGACCTGCTGCTGGAAGCATTTTCATGATAGAGAGAGCTCTCTCGG 905  
 DB 144633 TGTAGACCAAGACCTGCTGCTGGAAGCATTTTCATGATAGAGAGAGCTCTCTCGG 144574  
 QY 906 TGAACAGTCCAAATTAATATGATGTTTATATGAAGCCCAAGAGAGACCTTTGCCA 965  
 DB 144573 TGAACAGTCCAAATTAATATGATGTTTATATGAAGCCCAAGAGAGACCTTTGCCA 144514  
 QY 966 TGGCTAGTCTTCTCCATCCACCCCAACACTAATATATTTAGTGTGCTTTGTTA 1025  
 DB 144513 TGGCTAGTCTTCTCCATCCACCCCAACACTAATATATTTAGTGTGCTTTGTTA 144454  
 QY 1026 AAGCAAGATTAATCTTAACTGCTCTTACTCTTCCCTTAGCTAGCTAAATTAAGT 1085  
 DB 144453 AAGCAAGATTAATCTTAACTGCTCTTACTCTTCCCTTAGCTAGCTAAATTAAGT 144394  
 QY 1086 TGCATATAGCATTAATTAATTCGAGTCATGATGATCTCATGTTGATGATA 1145  
 DB 144393 TGCATATAGCATTAATTAATTCGAGTCATGATGATCTCATGTTGATGATA 144334  
 QY 1146 TTTTAACTAAGATATGATGATGATTTTTCAGAGTCCATTAATATTTATTTTC 1205  
 DB 144333 TTTTAACTAAGATATGATGATGATTTTTCAGAGTCCATTAATATTTATTTTC 144274  
 QY 1206 CTTTACTTTCTCACTCTGTTGAACAATTTGAAGATTTGGGAACCAATTTTGA 1265  
 DB 144273 CTTTACTTTCTCACTCTGTTGAACAATTTGAAGATTTGGGAACCAATTTTGA 144214  
 QY 1266 AACCAGATTCATATGATGATGATGATTTTTCATATTTCTGTTTGAAGATTTGG 1325  
 DB 144213 AACCAGATTCATATGATGATGATGATTTTTCATATTTCTGTTTGAAGATTTGG 144154  
 QY 1326 CCAATTAATAGTAATTTATATATGAGACTTGCCTGATACATTAATAGAGATTTTG 1385  
 DB 144153 CCAATTAATAGTAATTTATATATGAGACTTGCCTGATACATTAATAGAGATTTTG 144094  
 QY 1386 GACAAGAGATTTGCTGACAGCATTAATTAAGCTATTAAGCAATTAAGATATGAG 1445  
 DB 144093 GACAAGAGATTTGCTGACAGCATTAATTAAGCTATTAAGCAATTAAGATATGAG 144034  
 QY 1446 TGTGTTGACGTTTAATCATGATGATGATGATTTTCTATGATTTGAGAAATGTTTAA 1505  
 DB 144033 TGTGTTGACGTTTAATCATGATGATGATGATTTTCTATGATTTGAGAAATGTTTAA 143974  
 QY 1506 TATCAATCTAATTAATCTGAAATTTCTAAGAGCTTAATTTGCTTTGGCTGATGAGT 1565  
 DB 143973 TATCAATCTAATTAATCTGAAATTTCTAAGAGCTTAATTTGCTTTGGCTGATGAGT 143914  
 QY 1566 ATATTTGAATTTGTTGAATAATTAATTTCTATGTAATTAATTAATGCAAAAA 1625

DB 143913 ATATTTGAATTTGTTGAATAATTAATTTCTATGTAATTAATTAATGCAAAAA 143854  
 QY 1626 TATATTTGATGTTAAATCAATATAGATGATTTCTGTTTACATTTGTCATATGAATTAATATC 1685  
 DB 143853 TATATTTGATGTTAAATCAATATAGATGATTTCTGTTTACATTTGTCATATGAATTAATATC 143794  
 QY 1686 TGTGTTAATTTCAATTTGATATTAATTTGGCTTAAATATTTGATCTCAATTTATTTCTTC 1745  
 DB 143793 TGTGTTAATTTCAATTTGATATTAATTTGGCTTAAATATTTGATCTCAATTTATTTCTTC 143734  
 QY 1746 TCTGTTACTGTTAAATTAATATAGATGATGATTTTCTCAATTTGATCTCAATTTATTTCTTC 1805  
 DB 143733 TCTGTTACTGTTAAATTAATATAGATGATGATTTTCTCAATTTGATCTCAATTTATTTCTTC 143674  
 QY 1806 TATATTTAATTAATTAATTTTACTGTG 1832  
 DB 143673 TATATTTAATTAATTAATTTTACTGTG 143647

RESULT 4  
 AL672108/c  
 LOCUS 65229 bp DNA linear PRI 20-AUG-2002  
 DEFINITION Human DNA sequence from clone RP11-142117 on chromosome 10,  
 complete sequence.  
 ACCESSION AL672108  
 VERSION AL672108.7 GI:22449768  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 65229)  
 Clark, S.  
 Direct Submission  
 Submitted (20-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Aug 22, 2002 this sequence version replaced gi:22415876.  
 Genomes Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30).  
 An attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: EM: EMBL; SW:  
 SWISSPROT; TR: TrEMBL; WP: WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-142117 is from the library RPc1-11.1 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6.

FEATURES  
 source  
 1..65229  
 /organism="Homo sapiens"





/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyle ID No: 7476861c1"

BASE COUNT 565 a 378 c 438 g 533 t

ORIGIN

Query Match 31.5%; Score 576.4; DB 6; Length 1914;

Best Local Similarity 99.8%; Pred. No. 3.4e-83;

Matches 577; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

125 TTCACATCATGATATAGTGGCCCAATGAGAGACACGTTCTGCGATGTTCAACG 184

1337 TTCACATCATGATATAGTGGCCCAATGAGAGACACGTTCTGCGATGTTCAACG 1396

185 AAGGACGATATACCTTTTTCAGATATGCTTGAAGTCTTGGAACTCTGACT 244

1397 AAGGACGATATACCTTTTTCAGATATGCTTGAAGTCTTGGAACTCTGACT 1456

245 TTGGATTAAAGAAAGACTTCTGCTCTCCTGCTGAAATTCACAAAGTGGTTGCACCTC 304

1457 TTGGATTAAAGAAAGACTTCTGCTCTCCTGCTGAAATTCACAAAGTGGTTGCACCTC 1516

305 CTCATTAAGAACATGTTTGCACCTGCTGAAAGGCTTGGTATGATACATTCGCTTC 364

1517 CTCATTAAGAACATGTTTGCACCTGCTGAAAGGCTTGGTATGATACATTCGCTTC 1576

365 TTGGTTATAGTTATTTCTTCTTAAAGCTCCCTGAAAGGCAATATCATTTGGCTTG 424

1577 TTGGTTATAGTTATTTCTTCTTAAAGCTCCCTGAAAGGCAATATCATTTGGCTTG 1636

425 GGGTGATCACTGTTTACTTATGATCTTGTAGGCAATATCAAAATTAATCCACATTT 484

1637 GGGTGATCACTGTTTACTTATGATCTTGTAGGCAATATCAAAATTAATCCACATTT 1696

485 TCCAGTAAACAGATGTTTACATAAAGCATGTCAGCTTGGCTATTTGGTTGAAGGATTA 544

1697 TCCAGTAAACAGATGTTTACATAAAGCATGTCAGCTTGGCTATTTGGTTGAAGGATTA 1756

545 CAGAGCCCAATTAAGATTTAAATATTCATTAAGATTTTAAAGAGTGGCTGG 604

1757 CAGAGCCCAATTAAGATTTAAATATTCATTAAGATTTTAAAGAGTGGCTGG 1816

605 AGAGAGCTGAGATTTCCAGAGCTTGTAGTCTTATTCGCGAACAATTAAGCCCAAT 664

1817 AGAGAGCTGAGATTTCCAGAGCTTGTAGTCTTATTCGCGAACAATTAAGCCCAAT 1876

665 AATCATGACCTCTCCAGGCAATTTTAAAGACAGATGTC 702

1877 AATCATGACCTCTCCAGGCAATTTTAAAGACAGATGTC 1914

RESULT 6

AX302888 381 bp DNA linear PAT 30-NOV-2001

LOCUS AX302888 Sequence 208 from Patent WO01/9286.

DEFINITION AX302888

ACCESSION AX302888.1 GI:17383386

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 Jiang, Y., Dillon, D.C., Mitcham, J.L., Xu, J., Harlocker, S.L. and

Hepler, W.T.

Compositions and methods for the therapy and diagnosis of breast

cancer

Patent: WO 01/9286-A 208 25-OCT-2001;

CORIXA CORPORATION (US)

Location/Qualifiers

1. 381

/organism="Homo sapiens"

/db\_xref="taxon:9606"

BASE COUNT 106 a 68 c 51 g 142 t 14 others

ORIGIN

Query Match 19.7%; Score 360.6; DB 6; Length 381;

Best Local Similarity 95.3%; Pred. No. 2.2e-48;

Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

949 AGAGGAGACTTGTGGCATGCTGAGTCTTCTGATCCACCTTAACATTAATATTA 1008

1 AGAGGAGANNTTGGCATGCCGATGATNCTTCTATNCCACCTTAACATTAATATTA 60

1009 CTGATGCTGCTTGTAAAGCAATTAATCTTAACTTGGCTTCTTCTTGGCTT 1068

61 CTGATGCTGCTTGTAAAGCAATTAATCTTAACTTGGCTTCTTCTTGGCTT 120

1069 TGGTAACTAAATTAAGTGTATATAGCAATTAATTAATTAATTAATTAATTAAT 1128

121 TACCTAACTAAATTAAGTGTATATAGCAATTAATTAATTAATTAATTAATTAAT 180

1129 CTCATATGTTGATGATTTTTCACAACTAAGATCTATGATAGTTTTCACAGATTC 1188

181 CTCATATGTTGATGATTTTTCACAACTAAGATCTATGATAGTTTTCACAGATTC 240

1189 ATTAATATCTATTTATTCCTTACTTCTCACCCTGCTGGAACATTAAGAACTGAT 1248

241 ATTAATATCTATTTATTCCTTACTTCTCACCCTGCTGGAACATTAAGAACTGAT 300

1249 GGAACCAATTTTGGAAACAGATTCATAGTATGATGAATGAATGAATGAATGAAT 1308

301 GGAACCAATTTTGGAAACAGATTCATAGTATGATGAATGAATGAATGAATGAAT 360

1309 TTTTGAAGAAAGATGGCCAT 1329

361 TTTTGAAGAAAGATGGCCAT 381

RESULT 7

D64141 3090 bp mRNA linear ROD 13-FEB-1999

LOCUS D64141 Mus musculus mRNA for protein-tyrosine-phosphatase, complete cds.

DEFINITION D64141

ACCESSION D64141.1 GI:265457

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Mus musculus testis cDNA to mRNA.

protein-tyrosine-phosphatase.

1 (sites)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (sites)

Ohsugi, M., Kurumochi, S., Matsuda, S. and Yamamoto, T.

Molecular cloning and characterization of a novel cytoplasmic

protein-tyrosine phosphatase that is specifically expressed in

spermatocytes

J. Biol. Chem. 272 (52), 33092-33099 (1997)

2 (bases 1 to 3090)

Ohsugi, M.

Direct Submission

Submitted (15-SEP-1995) Miho Ohsugi, The University of Tokyo, 108,

Department of Oncology, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108,

Japan (E-mail: mhsugi@u-tokyo.ac.jp), Tel: 03-5449-5305,

Fax: 03-5449-5413

Location/Qualifiers

1. 3090

/organism="Mus musculus"

/db\_xref="taxon:10090"

/tissue-type="testis"

218. 1498

/ec\_number="3.1.3.48"

/codon\_start=1

/product="protein-tyrosine-phosphatase"

/protein\_id="BA03761.1"

/db\_xref="GI:265458"

/translation="MSSPRVVRKGTGRDNDDEBNSGNLNLRLNSLPSSQKMPRTKPI

ENKRNKSENVKPSHHSFSDKVELYYPELESDTDETVMDVSDRLBNRKNNDLQAK  
GSKTYSVPLSSSSRLSKDTETSSENELEQLQMOIRPLINSSRBSMRCLNTDQAK  
BELDIREFLELEQMTLPDDFNSGNTLQNDKNRDLIDYDSTRVLGNKNDITNS  
YRIYVHEEYEVIAFOGPEPIEDFOMVLEKNSVIMAMTRELICQYIKCSYMP  
ILKEBLEEHEEVSLEFETHVOYFTRVYQIWKNSGQCVHLEIDGCTGTA  
SADPEIKYRVYKSHITGPIYHCSAGVGTGVDVVFSAIEKNYSPDINIVT  
QNRKORGMIOREOYFCEIYEVLEVNLLALY\*

BASE COUNT 904 a 603 c 647 g 936 t  
ORIGIN

Query Match 16.8%; Score 308; DB 10; Length 3090;  
Best Local Similarity 59.3%; Pired. No. 3.3e-40;  
Matches 1079; Conservative 0; Mismatches 575; Indels 166; Gaps 26;

102 CAGATACGAGATATCTTCATTCATCATGATATAGTGGCCAAATGAGAGAAC 161  
1347 CTGGCAGGAGAGAACTACTCTTTTGAATGAAACATAGTACCCAGATGAGAAAGC 1406  
162 AACGCTCTGGCATGCTTCAACGAGAGAGATGATCTTTTGTACATATGCTGTG 221  
1407 AGGCTGTGGCATGATTCAAACGAGAGAGATGATCTTTTGTAAATGTGCTGTG 1466  
222 AAGTCTCTGGCACTGATCTGATTTGATTAAGAAAGCTGTGCTGCTCACTTGA 281  
1467 AAGTCTCTGAAACCTTCTGCTTTGATTAAGAGAGATCTGTG-CGCTGTCCCTGAG 1525  
282 ATTACCAAGTGGCTTGGACCTCTCTCATMAAGAACTGTTGACGTGCTGAGAGGCTT 341  
1526 GTTACCGAGAGAGCTTGGAGCC-----TGAGCCGCTGAGAAAGCG--TC 1565  
342 TGTATGACATACATCTGCTTCTTCTGTTATCATTTTCTTCTTAAAGCTCCT 401  
1566 TGGGGGCGGAGCTGCTGCTC-----TGATTTCTCTGAGAAAGCTCCT 1612  
402 GAAGGCAATATCATTTGGCTTGGGCTGATGATGTTACTTATGATCTGCTAGGCA 461  
1613 GAAGGTGACACTACTGGGACAGAGATGAGTACTGTTCCAC---TGATCTTTCTGACAA 1669  
462 TATCAAAATATCTCCCATATTTCCAGTGAACAGATGTTACATMAAAGCTGACGT 521  
1670 GACCAAAATACCTT-CGATGCTCTTACGGAAGGAGTGTGATGAAACCAACCTCCGT 1728  
522 TGGCTATTTGGTGAAGGATTAAGAGAGAGATTAAGATTAATATATCATTTAAG 581  
1729 TGGCTGCTGCTGTTGTTGATTTACAGAGCTTAATAAAGACTTACATCTCTTCA-TAAG 1787  
582 ATTTTATTTGAAAGTGGCTGGAGAGAGCTGAGATTTCCAGACTTGTAAAGTTCTTA 641  
1788 ATTTTATTTGAAAGTGGTGTAAAGAGCTGACAGACTTCTTACACTTTACAG-CCCTA 1846  
642 TTTGAGAGAGATTAAGGCAATAA--TCATGACCTTCCAGGCAATTTTAAAGACA 698  
1847 CTCTGGAGAGAGATGGGGGAGTAACTGACAGGCTCACAGACTGCCCTTGTGGATTT 1906  
699 T--GTCTATTCATTTCTTTAGCTAGAGCTGTACTTTTGTGGCATTTGATTAACCA 756  
1907 TTTGAGAGAGCTGAATGTGTGTAAGATCTATCTTCCAGCTGTGTAACGAGATAAC--- 1963  
757 GTTTAAAGAGAGCTGATTAAGTGGAGACTTAAGTGAACAAATTTGCTCATTTCTCT 816  
1964 --TCMAAGAGAGCTTAATTAAGGTGAGACTTTTGTATTAATTAAGGTTTGTGTCTT 2021  
817 ACATTTTCATA-----CTGCTGCCCTTCTCTAC 842  
2022 ATACTTCATGCTGCCAGAGAGCTGAGGATTTGAAGAGAGAGACTTACTCTCTGAGG 2081  
843 AGCTGCTAGACCAAGACTGTTGCTGTGGAGCATTTTATGATAGGAGAGAGCTCTC- 901  
2082 AGCAGTAGACCTTGTGCTGTGCTGCAAGCACTGTGAGTACTAGTGGCTCTCC 2141  
902 -----TCGGTGAACAGTCCAAACTAAATAGATTTTATATAGAA 942  
2142 TTAATATCCCTGAAACACATGCTGAGCAGTGCCTAAATTAATAGACATCTCTGTA 2201

943 AGCCAGAGAGAGACTTTTGGCATGCGCTGATCTTCTCATCCACCT-----AACACT 998  
2202 ATTC--AGTGAAGAGTGGTCCGGGCGGACGAGACTTCTCTCCTCCCTGCCACACT 2259  
999 TAACATATATCTACTGCTGCTTGTAAAGCAAGATTAACCTTAAGTCCCTTACT 1058  
2260 TAACATATCTTCCATGCTGCTTATCAGAGAGAGGCTTACACAGACAGGCTCTGAT 2319  
1059 CTTGCCCTTACATTAATTAAGTGTATATAGCATTTATATATATATCTGAGTCA 1118  
2320 CTTTACATGAGCCCAATTAATTAATGATGATTTG-----TTATTAAGTGTGATCA 2373  
1119 TTTGATGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1175  
2374 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2433  
1176 TT-----TCAGAGTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1230  
2434 TTTTCTCTTACAGAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2479  
1231 CATTTAGAAAGTGAATTTGGAGCCCAATTTTGGAAACCAAGTTCATGATGAAAT 1290  
2480 TATTTAGATGCTGAGATTTGG-ACATGATTTTGGAAAGCCCAATTAATTAATTAATTA 2358  
1291 GGAATCTTCATATCTGTTTGAAGAGATGAGCATTAATTAATTAATTAATTAATTA 1350  
2539 AAAAATTTTCCAGATGTTTGAAGAAACCTGCTGCTGCTTCCATTAATTAATTA 2596  
1351 GGACTTGTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1410  
2597 -----GAGCTGAGCAAAAT-GTGGTGTCTCTGTTGAAGCTTCAATGATTAAGCA 2649  
1411 TACCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1464  
2650 CTCCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2709  
1465 AATGAGATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1524  
2710 TGAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2769  
1525 AATTTCTAAGAGCTTATTTGCTTTTGGCTGAATGATTAATTAATTAATTAATTAAT 1584  
2770 TGGTTAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2829  
1585 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1643  
2830 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2889  
1644 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1692  
2890 AAGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2949  
1693 AATTTATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1752  
2950 ATATCAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3009  
1753 CTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1812  
3010 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3069  
1813 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1832  
3070 AAATGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3089

RESULT 8  
AC130106 173224 bp DNA linear HTG 08-AUG-2002  
LOCUS AC130106  
DEFINITION Rattus norvegicus clone CH230-17N23, \*\*\* SEQUENCING IN PROGRESS  
ACCESSION AC130106  
VERSION AC130106.1 GI:22138328

## KEYWORDS

## SOURCE

HTG: HTGS, PHASE1.  
Norway rat.

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

1 (bases 1 to 173224)

Munoz, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Albrooks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T., Barbata, J., Benton, J., Bimagne, K., Blankenburg, K., Bonin, D., Bouck, J., Bowles, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.E., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lora, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwa, S., Ogun, M., Okunola, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pul, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Sany, G., Scheraga, S., Scott, G., Shen, H., Shooshari, N., Slason, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Syarik, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, P., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinton, R., Wang, Q., Wang, S., Ward, Moore, S., Warren, R., Washington, C., Wallington, C., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

## TITLE

Unpublished

## JOURNAL

Direct Submission

## AUTHORS

Worley, K.C.

## REFERENCE

2 (bases 1 to 173224)

Submitted (08-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project information

Center project name: GWM

Center clone name: CH230-17N23

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap, version 0.990329

Consensus quality: 110803 bases at least Q40

Consensus quality: 120190 bases at least Q30

Consensus quality: 127171 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 77 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1639: contig of 1639 bp in length  
1640 1739: gap of unknown length  
1740 2901: contig of 1162 bp in length  
2902 3001: gap of unknown length  
3002 4294: contig of 1293 bp in length  
4295 4394: gap of unknown length  
4395 5609: contig of 1215 bp in length  
5610 5709: gap of unknown length  
5710 7058: contig of 1349 bp in length  
7059 7158: gap of unknown length  
7159 8627: contig of 1469 bp in length  
8628 8727: gap of unknown length  
8728 10289: contig of 1562 bp in length  
10290 10389: gap of unknown length  
10390 11570: contig of 1181 bp in length  
11571 11670: gap of unknown length  
11671 13440: contig of 1770 bp in length  
13441 13540: gap of unknown length  
13541 14876: contig of 1336 bp in length  
14877 14976: gap of unknown length  
14977 16738: contig of 1762 bp in length  
16739 16836: gap of unknown length  
16839 18011: contig of 1173 bp in length  
18012 18111: gap of unknown length  
18112 19712: contig of 1601 bp in length  
19713 19812: gap of unknown length  
19813 21882: contig of 2070 bp in length  
21883 21982: gap of unknown length  
21983 23866: contig of 1884 bp in length  
23867 23966: gap of unknown length  
23967 25663: contig of 1697 bp in length  
25664 25763: gap of unknown length  
25764 27201: contig of 1438 bp in length  
27202 27301: gap of unknown length  
27302 28420: contig of 1119 bp in length  
28421 28520: gap of unknown length  
28521 29588: contig of 1068 bp in length  
29589 29688: gap of unknown length  
29689 31206: contig of 1418 bp in length  
31207 31206: gap of unknown length  
31207 32829: contig of 1623 bp in length  
32829 32929: gap of unknown length  
32930 34640: contig of 1711 bp in length  
34641 34740: gap of unknown length  
34741 36598: contig of 1858 bp in length  
36599 36698: gap of unknown length  
36699 38197: contig of 1499 bp in length  
38197 38297: gap of unknown length  
38297 39479: contig of 1182 bp in length  
39480 39579: gap of unknown length  
39580 41047: contig of 1468 bp in length  
41048 41147: gap of unknown length  
41148 42801: contig of 1654 bp in length  
42802 42901: gap of unknown length  
42902 44190: contig of 1289 bp in length  
44191 44290: gap of unknown length  
44291 46297: contig of 2007 bp in length  
46298 46397: gap of unknown length  
46398 48473: contig of 2076 bp in length  
48474 48573: gap of unknown length  
48574 49797: contig of 1224 bp in length  
49798 49897: gap of unknown length  
49899 50906: contig of 1009 bp in length  
50907 51005: gap of unknown length  
51006 52842: contig of 1836 bp in length

[illegible]

QY	482	TTTTCCAGTGAACAGATGTTTCATATAAAACGATTTGAGCTTGCGATTTGGTTGGAAGGGA	541
Db	122468	CTTTCACATGTAAACTAACTTAACGTTACGTAAGAACAACTCAGCTTGCGTGTGGTGGTGA	122527
QY	542	TTACAGAGCCCAATTAAGAGATTAAATATATTCATTTAAGATTTTATTTTGAAGAGCTGC	601
Db	122528	TTACAGAGTTTAAATTAAGAGCTT-AAATGTGTTCGTATAGATTTTATTTGGAAGAGTGGC	122586
QY	602	TGAGAGAGCTGAGGATTTCCAGAGACTTTGTAAGTTCTTATTCGTGGAGACATAAGGCC	661
Db	122587	CAAAAGCCTCAGACATATTTCCAAACCTTTACAAAGC-CTGACTCTGGAGAACATGGGGGC	122645
QY	662	AATATCATGACCTCTCCAGGCATTTTAAAGACAGATCTATTCATGTCTCTTAGCTA	721
Db	122646	AATAAAC-TGACAGGCTCAGACAGCTACCTCTGTGGGAGTTAGACAGACATTTTATCT-	122703
QY	722	GAGCCTGACTTTTGTGCGCATTTGTAATAACCCAGATTAAAAAGATCCAGTTAGGGTG	781
Db	122704	GTGTCTAATACCTCCAGCTGTGTACACACACCTC-----AAAGAGCTCAATTAAGATG	122758
QY	782	GACTAAGCTTTGGACACAAATGGCTTCATTTCCATATTTTCATA-----	827
Db	122759	GACTAAGCTTTGATATTAATATGCGCTTTGTGTCTTATCTCCACAGTCCCAAGATGCC	122818
QY	828	-----CTGCTGCTTCCCTACAGCTGCTAGACCAAGAC	859
Db	122819	TGGGATTTGGAAGAGAGAGAGAGAGCTGATTCCTGAGAGCGACATGAGTCTCTGT	122878
QY	860	CTGTGGTGTGGGAAGCATTTTCATGGATAGGAGAGCTCCT-----CTCGGT	906
Db	122879	CTGCTGCTTGGGAAGCCCTTGTAGGTAAATGTGTCTCTTACCTGACACACACTGG	122933
QY	907	GAACAGTCCAAACTAAATAGATGTTTATAGAAAGCCAGAGAGAGACTTTTGCCAT	966
Db	122939	TGAACATCCCAAACTAAATTCACATCCATCCCTTAAGTC-AGTGGAAATGTTGGCAG	122996
QY	967	GCGTAGTTCTTCTCTATCCACCCCTAAC----ACTTAACATATTAAGTACTGCTGTGT	1023
Db	122997	ACCCAACTACACTTCTTCTCTATCCGCCCGAGTCAATGTACTCTCCACTGCGCTTAT	123056
QY	1024	TAAAGACAGATTTACCTTAAGTACTGCTCTTTCACCTTAAGTACATTAATAA	1083
Db	123057	CAGAAAGCATATTCATATACACTTCCCTCTCTATGTCT-AACTAGCCCTACATTTAAT	123115
QY	1084	GTTTGATATAGCATTATTAATTAATTCGAGTCATTCATGTATCTCTCATGTTGATG	1143
Db	123116	TTAATACATGATTAATGTCTGTGAATTTGAGTCATCTGTGAACGTGT-CTGTCAAG	123172
QY	1144	TATTTTCAACTAGATCTATGATAGTTTTTTTCAGAGTTCATTAATCATTTAT	1203
Db	123174	TTTTTTCTTTTAATAAGATCTAAAAAAGACTTTTTCCTTCGTGAGGACCA	123233
QY	1204	TGCTTACTCTCCACTGTTGGAACATTTGAACCTGAGATTGGGAAGCCCAATTTTG	1263
Db	123234	TTAAGCATTTCTTCTCTATAGAAATTTTGAATGCTGAGACAT- GAACCTGATTTTG	123292
QY	1264	GAAAAACAGATTCATAGTCATGAAATGGAACCTTCATATTCGTGTTTTGAAAGATGT	1323
Db	123293	AAGTACCAATTCATATGTTGTGAAATGGAACCTTCATATGTTGTAAAGAAA-TTCT	123355
QY	1324	GGCATTAATTACAGTAATTTTATTAATAGACTTGGCTCGTACAAATTAATAGTATTTT	1383
Db	123352	GGACATTTGTCATTAATCATAT--CAGATTCCTGCTCAGACAAAT- GTAGTGATTTCTT	123408
QY	1384	TGACACAGAGATTTGTGTGACAAAGCTATACCTATATATAGCTATAAACAATAGATATG	1443
Db	123409	TGGTGAAGATTTCAATGATCAGACAAATCTTAATTAAGCCCAAGAAATTAATGTAAG	123468
QY	1444	AGTGTTCAGAGT 1457	
Db	123469	TGTATTTGTACAGT 123482	

RESULT 9  
AC118715/c  
LOCUS AC118715 55785 bp DNA linear HTG 20-APR-2002  
DEFINITION Mus musculus clone RP24-14601, LOW-PASS SEQUENCE SAMPLING.  
AC118715  
AC118715.1 GI:20219209  
VERSION HTG: HTGS PHASE0.  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 55785)  
Biren, B., Linton, L., Nusbaum, C. and Lander, E.  
Mus musculus, clone RP24-14601  
Unpublished  
2 (bases 1 to 55785)  
Biren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Bouhassal, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Glade, S., Gort, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Huime, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., Melidim, J., Meneus, L.,  
Milnova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunhkhang, P., Plerre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppach, R.,  
Seaman, S., Sevel, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodores, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: 125107  
Center clone name: 146\_O\_1  
-----  
\* NOTE: This record contains 70 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1  
\* 708: contig of 708 bp in length  
\* 709 808: gap of 100 bp  
\* 809 1517: contig of 709 bp in length  
\* 1518 1617: gap of 100 bp  
\* 1618 2331: contig of 714 bp in length  
\* 2332 2431: gap of 100 bp  
\* 2432 3131: contig of 700 bp in length  
\* 3132 3231: gap of 100 bp

3322 3931: contig of 700 bp in length  
3932 4031: gap of 100 bp  
4032 4749: contig of 718 bp in length  
4750 4849: gap of 100 bp  
4850 5555: contig of 706 bp in length  
5556 5655: gap of 100 bp  
5656 6360: contig of 705 bp in length  
6361 6460: gap of 100 bp  
6461 7156: contig of 696 bp in length  
7157 7256: gap of 100 bp  
7257 7942: contig of 686 bp in length  
7943 8042: gap of 100 bp  
8043 8734: contig of 692 bp in length  
8735 8834: gap of 100 bp  
8835 9525: contig of 691 bp in length  
9526 9625: gap of 100 bp  
9626 10318: contig of 693 bp in length  
10319 10418: gap of 100 bp  
10419 11147: contig of 729 bp in length  
11148 11247: gap of 100 bp  
11248 11959: contig of 712 bp in length  
11960 12058: gap of 100 bp  
12060 12765: contig of 706 bp in length  
12766 12865: gap of 100 bp  
12866 13577: contig of 712 bp in length  
13578 13677: gap of 100 bp  
13678 14375: contig of 698 bp in length  
14376 14475: gap of 100 bp  
14476 15180: contig of 705 bp in length  
15181 15280: gap of 100 bp  
15281 15972: contig of 692 bp in length  
15973 16072: gap of 100 bp  
16073 16737: contig of 665 bp in length  
16738 16837: gap of 100 bp  
16838 17539: contig of 702 bp in length  
17540 17639: gap of 100 bp  
17640 18366: contig of 727 bp in length  
18367 18466: gap of 100 bp  
18467 19171: contig of 705 bp in length  
19172 19271: gap of 100 bp  
19272 19959: contig of 688 bp in length  
19960 20059: gap of 100 bp  
20060 20746: contig of 687 bp in length  
20747 20846: gap of 100 bp  
20847 21560: contig of 714 bp in length  
21561 21660: gap of 100 bp  
21661 22349: contig of 689 bp in length  
22350 22449: gap of 100 bp  
22450 23146: contig of 697 bp in length  
23147 23246: gap of 100 bp  
23247 23935: contig of 689 bp in length  
23936 24035: gap of 100 bp  
24036 24737: contig of 702 bp in length  
24738 24837: gap of 100 bp  
24838 25559: contig of 722 bp in length  
25560 25659: gap of 100 bp  
25660 26363: contig of 710 bp in length  
26370 26469: gap of 100 bp  
26470 27175: contig of 706 bp in length  
27176 27275: gap of 100 bp  
27276 27971: contig of 696 bp in length  
27972 28071: gap of 100 bp  
28072 28780: contig of 709 bp in length  
28781 28880: gap of 100 bp  
28881 29586: contig of 706 bp in length  
29587 29686: gap of 100 bp  
29687 30392: contig of 706 bp in length  
30393 30492: gap of 100 bp  
30493 31166: contig of 674 bp in length  
31167 31266: gap of 100 bp  
31267 31963: contig of 697 bp in length  
31964 32063: gap of 100 bp  
32064 32770: contig of 707 bp in length

```

* 32771 32870: gap of 100 bp
* 32871 33593: contig of 723 bp in length
* 33594 33693: gap of 100 bp
* 33694 34377: contig of 684 bp in length
* 34378 34477: gap of 100 bp
* 34478 35182: contig of 705 bp in length
* 35183 35282: gap of 100 bp
* 35283 35968: contig of 686 bp in length
* 35969 36068: gap of 100 bp
* 36069 36763: contig of 695 bp in length
* 36764 36863: gap of 100 bp
* 36864 37548: contig of 685 bp in length
* 37549 37648: gap of 100 bp
* 37649 38326: contig of 678 bp in length
* 38327 38426: gap of 100 bp
* 38427 39114: contig of 688 bp in length
* 39115 39214: gap of 100 bp
* 39215 39871: contig of 657 bp in length
* 39872 39971: gap of 100 bp
* 39972 40658: contig of 687 bp in length
* 40659 40758: gap of 100 bp
* 40759 41465: contig of 707 bp in length
* 41466 41565: gap of 100 bp
* 41566 42253: contig of 688 bp in length
* 42254 42353: gap of 100 bp
* 42354 43056: contig of 703 bp in length
* 43057 43156: gap of 100 bp
* 43157 43836: contig of 680 bp in length
* 43837 43936: gap of 100 bp
* 43937 44640: contig of 704 bp in length
* 44641 44740: gap of 100 bp
* 44741 45438: contig of 699 bp in length
* 45440 45539: gap of 100 bp
* 45540 46231: contig of 692 bp in length
* 46232 46331: gap of 100 bp
* 46332 47034: contig of 703 bp in length
* 47035 47134: gap of 100 bp
* 47135 47858: contig of 725 bp in length
* 47860 47959: gap of 100 bp
* 47960 48654: contig of 695 bp in length
* 48655 48754: gap of 100 bp
* 48755 49450: contig of 696 bp in length
* 49451 49550: gap of 100 bp
* 49551 50246: contig of 696 bp in length
* 50247 50346: gap of 100 bp
* 50347 51030: contig of 684 bp in length
* 51031 51130: gap of 100 bp
* 51131 51820: contig of 690 bp in length
* 51821 51920: gap of 100 bp
* 51921 52585: contig of 665 bp in length
* 52586 52685: gap of 100 bp
* 52686 53389: contig of 704 bp in length
* 53390 53489: gap of 100 bp
* 53490 54188: contig of 699 bp in length
* 54189 54288: gap of 100 bp
* 54289 54977: contig of 689 bp in length
* 54978 55077: gap of 100 bp

```

Query Match Best Local Similarity 5.8%; Score 106.6; DB 2; Length 55785; Pred. No. 3e-08; Matches 308; Conservative 0; Mismatches 134; Indels 41; Gaps 8;

```

QY 186 AGGAGCAGTACTTTGTTGATGATGCTTGAAGTTCTTGGAACTTGTGACTT 245
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44434 AGGAGCAGTACTTTGTTGATGATGCTTGAAGTTCTTGGAACTTGTGACTT 44375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 246 TGGATTAGAAAGACTCTGTGCTCTCATCTGAAATTACCAAGTGGTTGACCTCC 305
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44374 CGATTAGAGAGACTGTG-CGCTGTGCTCCCGAGGTTACCGACAGCTTGGAGCC--- 44319
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 306 TCATTAAGAACATGTTGCTGCTGCTGAGAGGCTTTCTATGATACATCATCTGCTTCT 365
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44318 -----TGAGCGCTGCTGAAGG-TCGCGGCGCGTGCAGCTGTGCTTCT 44276
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

QY 366 TGGTTATGATTTATTTCTTTCTAAAGCTCCGTAAGGCAATATCATTTGGCTGG 425
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44275 GATTTT-----TCTCTGTAAGTCCCTGAGGTAAGCACTACTGGGACAG 44230
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 426 GGTGATCACTGTTACTTATTTGATCTTGTGCTAGCAATATCAATTAATCTCCGATTTT 485
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44229 ACTGAACGTCTTCCAC---TTGATCTTCTGACAAAGAGCAAAATACCC-CCATGCTT 44174
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 486 CCAGTGAACAGATGTTACATTAAGCAATGACGTGCTGCTATTTGGTTAAGGATTAAC 545
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44173 CTACGGAACGGAAGTGCATGAACAACCTCCGCTGCTGTGTTGTGTATTAAC 44114
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 546 AGAGCCCATTAAGATTTAAATATATCATTAAGATTTATTTGGAAGGTGGCTGGA 605
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44113 AGAGCTTAATTAAGACTT-AGATGTGTTGATTAAGATTTATTTGGAAGGTGGTAAA 44055
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 606 GAGAGCTGAGATTTCCAGACCTTGTGAATGTTCTATTTGGAAGCAATAGCCATA 665
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44054 AGCCTGACAAATTTCTTAACACTTCAAG-CCCTACTGTGGGAAACATGGGGCAGTA 43996
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 666 ATC 668
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 43995 AAC 43993
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 10
AC073269 137889 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-436F9 from 7, complete sequence.
DEFINITION AC073269
AC073269 GI:14269824
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 137889)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
2 (bases 1 to 137889)
VanBrunt, A., Abbott, A. and McEellan, M.
The sequence of Homo sapiens BAC clone RP11-436F9
Unpublished (2001)
3 (bases 1 to 137889)
Waterston, R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 137889)
Waterston, R.H.
Direct Submission
Submitted (31-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 137889)
Waterston, R.H.
Direct Submission
Submitted (01-JUN-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 137889)
Waterston, R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 137889)
Waterston, R.
Direct Submission

```

JOURNAL  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On May 31, 2001 this sequence version replaced g1:13794256.

COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0436P09

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send  
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The Rpt1-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at the Roswell Park Cancer Institute  
(<http://bacpac.med.buffalo.edu>)  
VECTOR: pBAC3.6  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is Rp11-29C24, 2000 bp overlap; the  
clone sequenced to the right is CTD-2016H9, 200 bp overlap. Actual  
start of this clone is at base position 1 of Rp11-436F9; actual end  
is at base position 92636 of CTD-2016H9.

Polymorphisms have been identified between AC069313, AC021448 and  
AC073269.

FEATURES  
Source

Location/Qualifiers  
1. 137889  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7"  
/clone="Rp11-436F9"  
/clone\_lib="Rpt1-11"  
453..302  
repeat\_region  
/rpt\_family="AT-rich"  
774..1051  
repeat\_region  
/rpt\_family="Alu"  
1063..1316  
repeat\_region  
/rpt\_family="L1"  
1351..1469  
repeat\_region  
/rpt\_family="L1"  
2098..2256  
repeat\_region  
/rpt\_family="MIR"  
2505..2753  
repeat\_region  
/rpt\_family="L1"

repeat\_region 2971..3034  
/rpt\_family="L1"  
repeat\_region 3558..3742  
/rpt\_family="L2"  
repeat\_region 3905..4033  
/rpt\_family="L2"  
repeat\_region 4113..5347  
/rpt\_family="L1"  
repeat\_region 4854..4888  
/rpt\_family="CT-rich"  
5341..5560  
/rpt\_family="Alu"  
6804..6903  
/rpt\_family="L2"  
6920..7957  
/rpt\_family="MER1\_type"  
7958..8260  
/rpt\_family="Alu"  
8261..8370  
/rpt\_family="MER1\_type"  
8826..8873  
/rpt\_family="ERV1"  
8833..8877  
/rpt\_family="(TG)n"  
8836..9211  
/rpt\_family="L2"  
9345..9387  
/rpt\_family="L2"  
10487..10723  
/rpt\_family="MER2\_type"  
13743..14036  
/rpt\_family="Alu"  
14027..14069  
/rpt\_family="(TAAA)n"  
14540..14937  
/note="match to EST AV729823 (NID:g10839244)"  
14800..14872  
/note="match to EST AV709400 (NID:g10726999)"  
14800..14872  
/note="match to EST BP966467 (NID:g12333682)"  
14824..14872  
/note="match to EST A1138808 (NID:g3644780) qd98g02.x1"  
14918..15301  
/rpt\_family="ERV1"  
15247..15529  
/rpt\_family="(TCTA)n"  
15308..15546  
/rpt\_family="ERV1"  
15566..15748  
/rpt\_family="(TCTA)n"  
15626..15894  
/rpt\_family="ERV1"  
15764..15802  
/rpt\_family="(TCTA)n"  
15814..15980  
/rpt\_family="(TCTA)n"  
15990..16035  
/rpt\_family="ERV1"  
16102..16395  
/rpt\_family="MALR"  
16405..17052  
/note="match to EST BP966467 (NID:g12333682)"  
16872..17339  
/note="match to EST AV709400 (NID:g10726999)"  
16872..17204  
/note="match to EST A1138808 (NID:g3644780) qd98g02.x1"  
17326..17347  
/rpt\_family="AT-rich"  
17664..17970  
/rpt\_family="Alu"  
17946..17977  
/rpt\_family="A-rich"  
18654..18690

repeat\_region /rpt\_family="MIR"  
18709..18849  
repeat\_region /rpt\_family="L2"  
19230..19400  
repeat\_region /rpt\_family="(TATATC)n"  
19476..19668  
repeat\_region /rpt\_family="Alu"  
19850..20013  
repeat\_region /rpt\_family="L2"

Query Match 5.3%; Score 97.6; DB 9; Length 137889;  
Best Local Similarity 47.4%; Pred. No. 6.4e-07;  
Matches 361; Conservative 0; Mismatches 394; Indels 7; Gaps 2;

1073 TAACTAATAAGTTGATATAGCATTTATTAATTTCTGACATTCAGTCTCT 1132  
1133 CATGTTGATGTTTTCCTCAACTAGATGATGATTTTTCAGAGTTCATTA 1192  
63447 ATTTTATTAATTAATTAATTAATTAATTTTATTAATTAATTAATTTT 63506  
1193 AATCATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1252  
63507 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 63566  
1253 ACCCAATTTGGAAACCAAGTTCATGATGATGATGATGATGATGATGAT 1312  
63567 TTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 63626  
1313 TGAAGAATGAGCATTAATTAATTAATTAATTAATTAATTAATTAATTA 1372  
63627 TTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 63682  
1373 TAGATTAATTTTGGACAGAGGCTTGTGACACATTAATTAATTAATTA 1432  
63683 TATTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 63742  
1433 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1492  
63743 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 63802  
1493 AGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1552  
63803 TTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 63862  
1553 TGCTGAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1612  
63863 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 63922  
1613 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1669  
63923 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 63982  
1670 CATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1729  
63983 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 64042  
1730 CTAAATTAATTTCTCTGTTACTGTAATAATTAATTAATTAATTAATTA 1789  
64043 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 64102  
1790 CAGAAGAAATCTGATCTATTAATTAATTAATTAATTAATTAATTAAT 1831  
64103 CATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 64144

RESULT 11  
AL603966/c 103737 bp DNA 1linear PRI 30-JAN-2002  
LOCUS AL603966 Human DNA sequence from clone Rpl1-164N7 on chromosome 10, complete  
DEFINITION sequence.  
ACCESSION AL603966

VERSION AL603966.9 GI:18476754  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 103737)  
BIRD.C.  
TITLE Direct Submission  
JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
On Feb 1, 2002 this sequence version replaced gi:16605768.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Mp., MOPREP; Information on the MOPREP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormrep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/MOP/Chr10  
Rpl1-164N7 is from the library RPlC1-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: PBACe3.6

## COMMENT

IMPORTANT: This sequence is not the entire insert of clone  
Rpl1-164N7. It may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true left end of clone Rpl1-164N7 is at 1 in this sequence. The  
true left end of clone Rpl1-13E1 is at 101738 in this sequence.  
FEATURES  
source  
1..103737  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="Rpl1-164N7"  
/clone\_11D="RPlC1-11.1"  
BASE COUNT 31673 a 22545 c 21304 g 28215 t

## BASE COUNT

Query Match 5.2%; Score 95; DB 9; Length 103737;  
Best Local Similarity 95.1%; Pred. No. 1.8e-06;  
Matches 98; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

28 TTTATGCTTTAGAACTTAAGATCTGCTGTGATGATCACTGGGAATCAACCAAG 87  
1133 CATGTTGATGTTTTCCTCAACTAGATGATGATTTTTCAGAGTTCATTA 1192  
63447 ATTTTATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTTT 63506  
1193 AATCATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1252  
63507 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 63566  
1253 ACCCAATTTGGAAACCAAGTTCATGATGATGATGATGATGATGATGAT 1312  
63567 TTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 63626  
1313 TGAAGAATGAGCATTAATTAATTAATTAATTAATTAATTAATTAATTA 1372  
63627 TTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 63682  
1373 TAGATTAATTTTGGACAGAGGCTTGTGACACATTAATTAATTAATTA 1432  
63683 TATTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 63742  
1433 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1492  
63743 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 63802  
1493 AGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1552  
63803 TTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 63862  
1553 TGCTGAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1612  
63863 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 63922  
1613 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1669  
63923 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 63982  
1670 CATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1729  
63983 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 64042  
1730 CTAAATTAATTTCTCTGTTACTGTAATAATTAATTAATTAATTAATTAAT 1789  
64043 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 64102  
1790 CAGAAGAAATCTGATCTATTAATTAATTAATTAATTAATTAATTAAT 1831  
64103 CATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 64144

RESULT 12  
AL450334/c 171930 bp DNA 1linear PRI 21-JUN-2002  
LOCUS AL450334 Human DNA sequence from clone Rpl1-556L1 on chromosome 10, complete  
DEFINITION sequence.  
ACCESSION AL450334



```

VERSION      AL450334.15   GI:21540028
KEYWORDS     HTG.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE    Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS      Heath, P.
TITLE        Direct SubMISSION
JOURNAL       Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
COMMENT       Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
              On Jun 23, 2002 this sequence version replaced gi:14018292.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence was finished as follows unless otherwise noted: all
              regions were either double-stranded or sequenced with an alternate
              chemistry or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by at least
              one plasmid subclone or more than one M13 subclone; and the
              assembly was confirmed by restriction digest. The following
              abbreviations are used to associate primary accession numbers given
              in the feature table with their source databases: Em:, EMBL; Sw:,
              SWISSPROT; Tr:, TREMBL; Wp:, WormPeP; Information on the WormPeP
              database can be found at:
              http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
              was generated from part of bacterial clone contigs of human
              chromosome 10, constructed by the Sanger Centre Chromosome 10
              Mapping Group. Further information can be found at
              http://www.sanger.ac.uk/MGP/Chr10
              RPL1-556L1 is from the library RPci-11.2 constructed by the group
              of Pieter de Jong. For further details see
              http://www.chori.org/bacpac/home.htm
              VECTOR: pBACE3.6
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquery@sanger.ac.uk
              -----
FEATURES             Location/Qualifiers
     source           1..171930
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RPL1-556L1"
                     /clone_1lb="RPCT-11.2"
BASE COUNT          54090 a 36289 c 33202 g 48349 t
ORIGIN
Query Match       5.2%; Score 95; DB 9; Length 171930;
Best Local Similarity 95.1%; Pred.No.1.6e-06;
Matches 98; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
1y               28 TTTATGGCTTTAGAACTTAGAATTCGCCTGTGAGTCAACTCTGGGAATCAACCAAGC 87
                |||
2b               2184 TTAAAGCGTTTAGAACTTAAGAATTCGCTGTGAGTCAACTCTGGGAATCAACCAAGC 2125
1y               88 AACAGAGAAAAAACAGATACCAGGATTTCTTCATCATTCAC 130
                |||||
2b               2124 AACAGAGAAAAAACAGATACCAGGATATTCTTCATCATTAAGC 2082

RESULT 13
LOCUS         AC069435               111861 bp     DNA                    linear            PRI 30-APR-2002
DEFINITION    Homo sapiens 3 BAC RPL1-384L1 (Roswell Park Cancer Institute Human
              BAC Library) complete sequence.
```

AC069435  
AC069435.19 GI:20340410  
HTG.  
Homo sapiens.  
Homo sapiens.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 11861)  
Mizny,D.M., Adams,C., Adlo-Oduola,B., Ali-oman,F.R., Allen,C., Alshrooks,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,J., Benton,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bouch,J., Bowie,S., Brivaga,M., Brown,E., Brown,M., Bryant,N.P., Buahy,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douchnaite,R.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flaigg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs1,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyvet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,M., Lousgheed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mathewney,E., Mcleod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohapat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nnokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oyedele,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Plinius,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitaishvili,N., Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,N., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Naylor,S.L., Weinstein,G. and Gibbs,R.

Direct Submission  
Unpublished  
2 (bases 1 to 11861)  
Worley,K.C.  
Direct Submission  
Submitted (30-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 11861)  
Worley,K.C.  
Direct Submission  
Submitted (30-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
ON Apr 30, 2002 this sequence version replaced gi:20066162.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.



	AC005504	104992 bp	DNA	linear	HMC 01-APR-1999
	Plasmodium falci-parum chromosome 12, *** SEQUENCING IN PROGRESS				
	***, 3 unordered pieces.				
AC005504					
AC005504.3	GI:458584				
HMG: HTGS_PHASE1					
Plasmodium falci-parum.					
Plasmodium falci-parum					
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
1 (bases 1 to 104992)					
Hyman, R.W., Fung, E.L., Qin, F., Tanaki, T., Kurd, O.B., Conway, A.B. and Davis, R.W.					
Plasmodium falci-parum 3D7 chromosome 12					
Unpublished					
2 (bases 1 to 104992)					
Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.					
Direct Submission					
Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA					
On Apr 2, 1999 this sequence version replaced gi:4337172.					
* NOTE: This is a "working draft" sequence. It currently					
* consists of 3 contigs. The true order of the pieces					
* is not known and their order in this sequence record is					
* arbitrary. Gaps between the contigs are represented as					
* runs of N, but the exact sizes of the gaps are unknown.					
* This record will be updated with the finished sequence					
* as soon as it is available and the accession number will					
* be preserved.					
1	58642:	contig of 58642 bp in length			
*	58842:	gap of unknown length			
*	58843	91011: contig of 32169 bp in length			
*	91012	91211: gap of unknown length			
*	91212	104992: contig of 13781 bp in length.			
Location/Qualifiers					

[illegible]

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
Plasmodium falciparum	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	1 (bases 1 to 169546)	Hyman, R.W., Fung, E.L., Qin, F., Rowley, D., Mao, J., Tamaki, T., Kurd, O.B., Conway, A.B. and Davis, R.W.	Plasmodium falciparum 3D7 chromosome 12
unpublished	2 (bases 1 to 169546)	Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.	Direct Submission	Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology